# Chemicals Catalog Databases: an Overview and Evaluation

Martin Paul Brändle*, Engelbert Zass

Informationszentrum Chemie Biologie Pharmazie, ETH Zurich, 8093 Zurich, Switzerland

## Abstract

In addition to the time-honored printed catalogs, many different electronic information sources have become available for the procurement of chemicals. Examples are Web catalogs of individual suppliers, specialized search engines such as eMolecules, and well-established catalog databases such as ACD (Available Chemicals Directory), Chem Sources, or CHEMCATS. This manifold of sources is complicated due to a variety of different interfaces, providers, and database versions. We present an overview on important sources and the corresponding interfaces, their availability, and the meta information provided. Finally, search results for some examples by using major sources are compared.

## 1 Introduction

What Amazon [1] is for book aficionados, are chemicals catalog databases for chemists, biologists and pharmaceutical scientists working in academia and industry: They aggregate printed or electronic product information from various suppliers into a single, searchable interface and disburden the user from searching multiple catalogs for the products needed in his research. But are they like Amazon? Chemicals are not books; there are laws and safety regulations that restrict the market to approved professionals and complicate business. Like books, however, they can be searched by name, and analogous to the ISBN for books, they have identification numbers such as CAS Registry Numbers (CAS RN) or EC numbers in the European Union. In addition, they can be searched by structure or properties. Well-established commercial catalog databases such as ACD (Available Chemicals Directory), Chem Sources, or CHEMCATS offer search facilities for these properties. Recently, they face competitors that are freely available on the WWW. The Web has enabled creation of a new generation of chemicals directories that have different purposes: search engines for chemicals, portals which among chemicals also allow one to find other product categories, and specialized databases for screening compounds that are commercially available. In this report, we present an evaluation of the features of the major commercial and selected free sources, and compare results of test searches we had carried out.

For evaluation, we considered the following criteria grouped in categories: 1) Content. How many unique compounds, products and suppliers are contained, are product prices provided, what are the update frequency and the growth rate, classes of compounds covered, data fields available, are there any extra data such as physical data, safety information, regulations, links to literature? 2) Access for commercial databases. How can the database be accessed, is it part of a larger system, what are price and licensing models? 3) Technical requirements. Which platforms are supported, is installation easy, which software and plugins are used, in which browsers does product work? 4) Search facilities: Which fields can be searched, what search types (structure, substructure, similarity etc.) can be used, is import of structures and data possible, can operations be performed in batch mode? 5) Special functions. Are links to external data provided, can quotes be requested, can products be ordered directly with an included (web-)shop or indirectly by sending orders by e-mail or fax, what import and export facilities are available, can the database be integrated in chemical information managements (CIM) systems for inventory and process management, are current awareness functions for tracking new and updated products available? 6) User interface. Is the search mask concise, is the structure editor user-friendly and does it allow for complex structures such as organometallics, is display of hits clear, is record navigation available, how many screens have to be passed, are functions such as search history and sorting provided? 7) User support. Does the documentation describe the product adequately and help the user in case of problems?

In Section 2 we describe major sources for chemicals and apply these criteria. Section 3 compares results of searches we had carried out for about 20 compound types in commercial and free databases. Discussion and conclusions follow in Sections 4 and 5, respectively.

## 2 Sources

Table 1a allows comparison of the features and provides our assessment of the major commercial catalog databases CHEMCATS, ChemACX, Available Chemicals Directory, and Chem Sources. Table 1b provides this for the free sources ChemExper, ChemNet, eMolecules, ICISsearch, LabVelocity, Super Natural Database, and ZINC. In the following, a few of these product features will be discussed in detail.

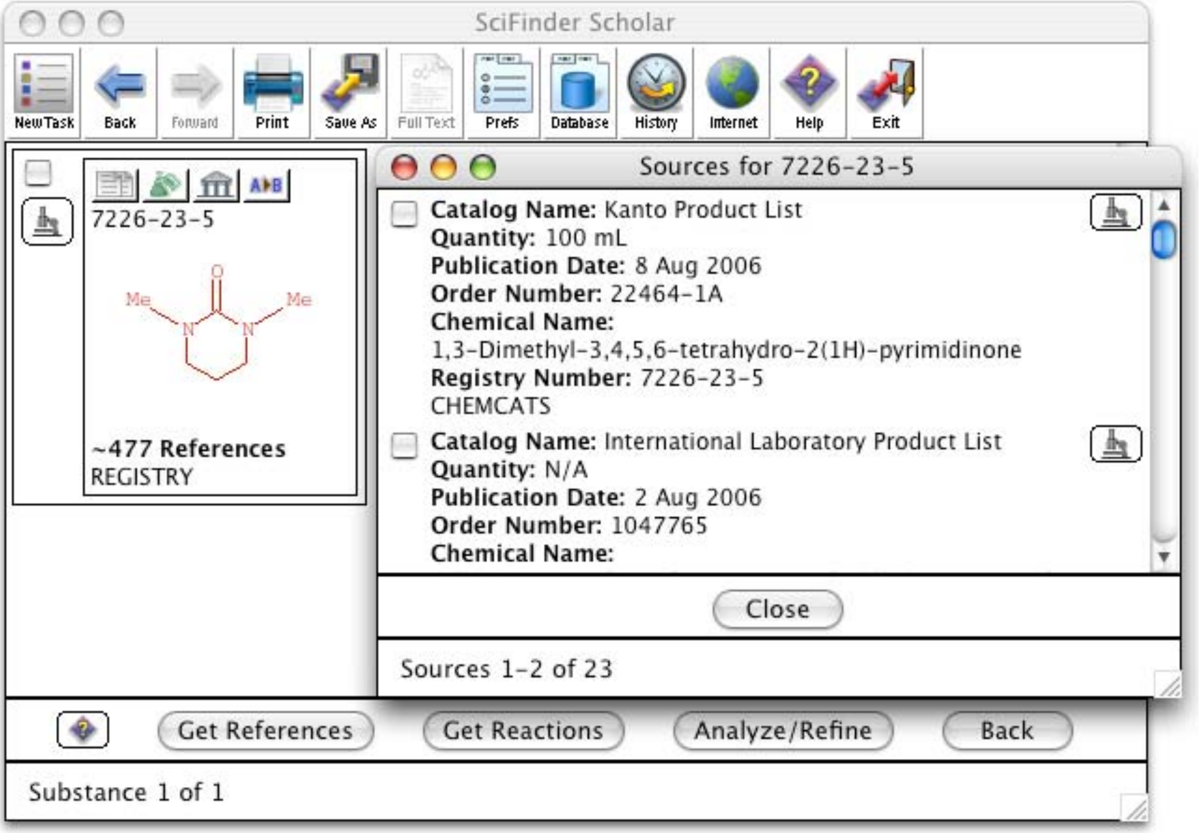
### 2.1 CHEMCATS

CHEMCATS (Chemical Catalogs Online) is produced by Chemicals Abstract Service (CAS) and comprises more than 11 million product records on chemicals, enzymes, proteins and other biochemicals [2,3]. Our searches revealed that it contains more than 3.4 million compounds as registered in CAS Registry. This puts it on top rank as the largest chemicals catalog database. The data is obtained from over 800 suppliers and over 900 catalogs (coverage 1995-). The database can be searched through the SciFinder, SciFinder Scholar, STN (STN Express, command-line driven STN Messenger language), STN on the Web, and STN Easy interfaces. Searches can be effected directly in STN CHEMCATS on chemical and trade names, company names and addresses, but not on structures. These must first be searched in CAS REGISTRY, and then the resulting CAS Registry Numbers (CAS RNs) transfered to CHEMCATS. Sorting of the CHEMCATS results is possible for fields catalog and company name, linear structural formula, occurrence of hit terms, order number, publication date and year, but not for price. As with most host databases, pricing of CHEMCATS on STN is pay-per-use.

A graphical user interface to CHEMCATS is provided by SciFinder Scholar and SciFinder, see Figure 1. Both provide also access to the CAS databases CAplus, CAS Registry, CASREACT, CHEMLIST, and the NIH database Medline. SciFinder, which is the software version available for the chemical industry, offers more options (e.g. alerts, analysis functions and visualization) than the academic Scholar version. The SciFinder (Scholar) client software must be installed locally on a Windows PC or an a Apple Mac computer with OS X. Installation is straightforward; however, our experience shows that Mac OS X users sometimes have problems to place the required site license file at the correct location. There are different purchasing options: For the industrial SciFinder version, subscription pricing or task packages where block of tasks can be bought are available, whereas for academia, a subscription model that depends on the number of seats (concurrent users) and the type of institution (university, college) is offered. Our test searches were executed in SciFinder Scholar and STN CHEMCATS.

The SciFinder (Scholar) interface was designed to guide a user along a pathway, called „task“ in SciFinder terminology. Within a chosen task, users can move forth and back along the task steps. Until SciFinder version 2006, results of different tasks can not be combined; this will change with version 2007. In contrast to STN, access to CHEMCATS records is possible via CAS Registry records only. Three main pathways are available to find CAS Registry entries: Explore by Substance (exact structure, substructure and similarity search), Explore by Molecular Formula, and Locate by Substance Identifier (chemical name, CAS RN). Whereas the first pathway allows to restrict the search to commercially available substances from the beginning (through the Filters option in the Get Substances dialog window), the latter two show all compounds found in the CAS Registry whether commercially available or not. An additional refinement step has to be carried out to get only the trade chemicals. The display of structure hits shows a flask button if there are CHEMCATS entries (Figure 1), which upon clicking yields a list of all available sources. No sort order is apparent for this data, and sorting is not possible. Data can be exported as plain ASCII, quoted (comma-separated), Rich Text, and tagged format. Individual entries can be cherry-picked for printing. Pricing and company address information is obtained by clicking on the microscope button for a product record. Prices, however, are sometimes missing and must then be requested from the supplier.

**Figure 1**. Screenshot of SciFinder Scholar showing CHEMCATS sources for DMPU. SciFinder Scholar™ screens are reproduced with the permission of CAS, a division of the American Chemical Society



7226-23-5

~477 References
REGISTRY

□ **Catalog Name:** Kanto Product List
**Quantity:** 100 mL
**Publication Date:** 8 Aug 2006
**Order Number:** 22464-1A
**Chemical Name:**
1,3-Dimethyl-3,4,5,6-tetrahydro-2(1H)-pyrimidinone
**Registry Number:** 7226-23-5
CHEMCATS

□ **Catalog Name:** International Laboratory Product List
**Quantity:** N/A
**Publication Date:** 2 Aug 2006
**Order Number:** 1047765
**Chemical Name:**

Sources 1-2 of 23

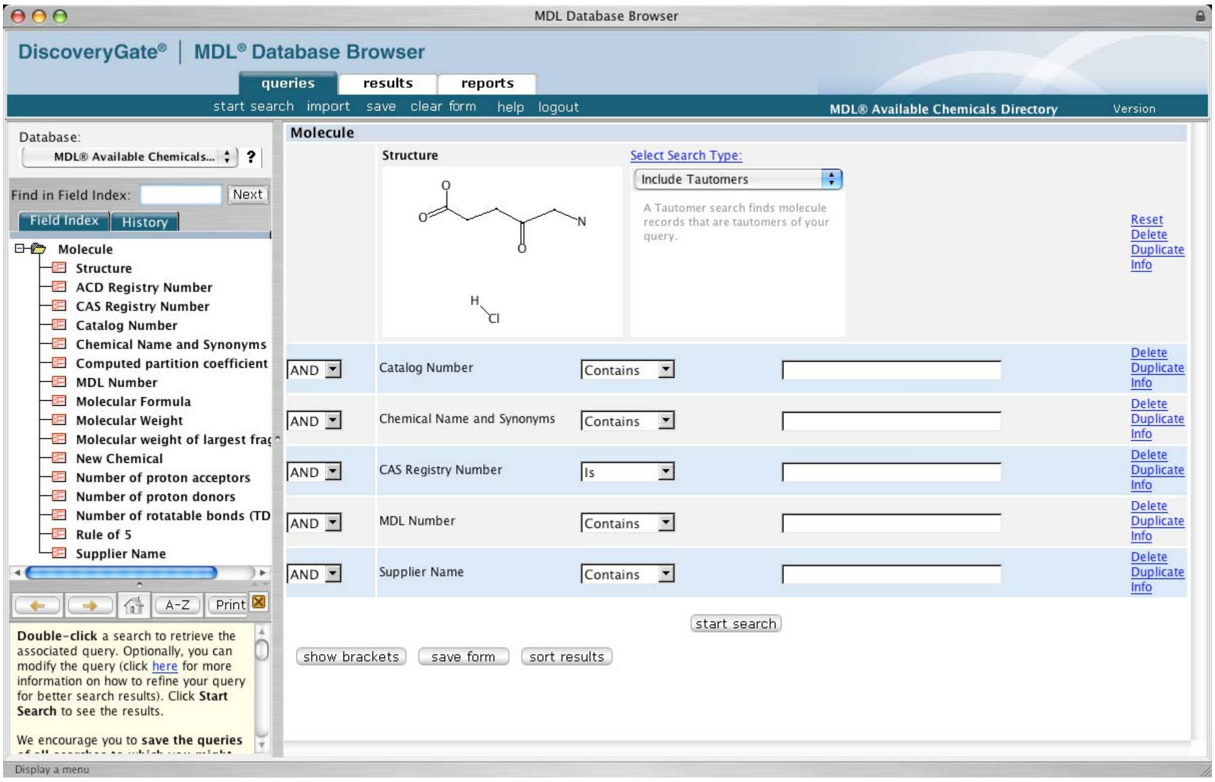
Substance 1 of 1

## 2.2 ACD (Available Chemicals Directory)

The Available Chemicals Directory may be called "the mother of all catalog databases", as Fraser-Williams already in 1988 created this database from the printed Fine Chemicals Directory which originated in 1981, based on an earlier version, the CAOCI Index to Commercially Available Chemicals (February 1979) [4]. It is now produced by MDL, and offered in the well-established client-server database systems ISIS, ISENTRIS, and recently also under the new Web interface DiscoveryGate [5]. These diverse systems provide not only powerful retrieval options for the ACD, they also integrate this catalog database into an impressive and useful array of other compound and reaction databases available under the same retrieval systems, as well as in chemical information management systems (CIM) for inventories and process management. ACD as one of the largest catalog databases provides access to 1.4 million products from over 680 catalogs, representing 510'000 compounds. The information is updated quarterly; until 2004, updates were biannually. As a special feature, the ACD also contains calculated 3D structures for use in modelling.

Like other vendors, MDL has been producing since 1997 a special catalog database for screening compounds: ACD-SD, now MDL Screening Compounds Directory, with about 6 million products (3.4 mio. structures) from 46 suppliers specialized on this type of products. In DiscoveryGate, these may be searched alone or together with the ACD.

**Figure 2.** Screenshot of MDL DiscoveryGate showing a query for δ-aminolevulinic acid hydrochloride in ACD. Copyright © 2006 MDL Information Systems Inc.



Database: MDL® Available Chemicals...

Molecule
- Structure
- ACD Registry Number
- CAS Registry Number
- Catalog Number
- Chemical Name and Synonyms
- Computed partition coefficient
- MDL Number
- Molecular Formula
- Molecular Weight
- Molecular weight of largest frag
- New Chemical
- Number of proton acceptors
- Number of proton donors
- Number of rotatable bonds (TD
- Rule of 5
- Supplier Name

**Double-click** a search to retrieve the associated query. Optionally, you can modify the query (click here for more information on how to refine your query for better search results). Click **Start Search** to see the results.

We encourage you to **save the queries**

Molecule

Structure

Select Search Type: Include Tautomers

A Tautomer search finds molecule records that are tautomers of your query.

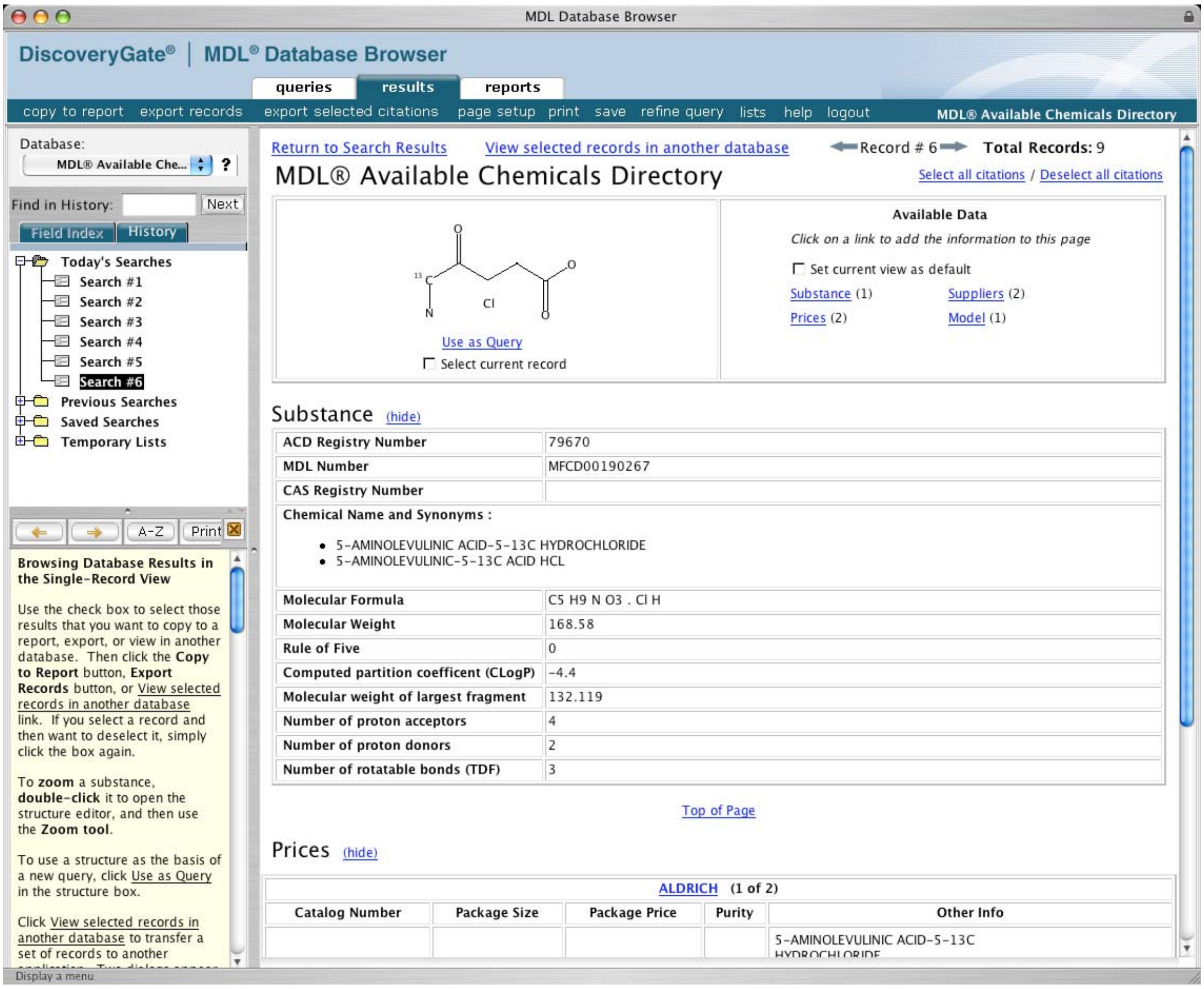
| | | | |
|---|---|---|---|
| AND | Catalog Number | Contains | |
| AND | Chemical Name and Synonyms | Contains | |
| AND | CAS Registry Number | Is | |
| AND | MDL Number | Contains | |
| AND | Supplier Name | Contains | |

For our comparison, we used the ACD under the interface (retrieval system) DiscoveryGate [5]. Figure 2 shows a structure search for δ-aminolevulinic acid hydrochlorides. The chosen search option “Include Tautomers” leads to retrieval also of isotope-labelled compounds with that structure, our main focus in this search (cf. Table 5). Other search fields shown in the screen shot like CAS RN are empty, and will thus be ignored. With appropriate input, they may be combined with Boolean logic, including parentheses („show brackets“) for nesting. The search mode in text fields can be set to „is“ (exact match), „contains“, „begins with“,

„ends with“, and „exists“. This permits very flexible searches for complete names as well as for name fragments, e.g., („contains“ PEG OR „contains“ polyethyleneglycol) AND „contains“ 8000.

The search mask may be tailored by using the „Duplicate“ or „Delete“ links (at the right of Figure 2), as well as built from scratch using the searchable tree view of fields displayed in the upper left frame of the screen. Tailored masks may be saved for further use. The help messages in the lower left frame are quite useful, but may be turned off any time by experienced users.

**Figure 3.** Screenshot of MDL DiscoveryGate showing ACD sources for 13C-labelled δ-aminolevulinic acid hydrochloride. Copyright © 2006 MDL Information Systems Inc.



**Browsing Database Results in the Single-Record View**

Use the check box to select those results that you want to copy to a report, export, or view in another database. Then click the **Copy to Report** button, **Export Records** button, or View selected records in another database link. If you select a record and then want to deselect it, simply click the box again.

To **zoom** a substance, **double-click** it to open the structure editor, and then use the **Zoom tool**.

To use a structure as the basis of a new query, click Use as Query in the structure box.

Click View selected records in another database to transfer a set of records to another

MDL® Available Chemicals Directory

*Click on a link to add the information to this page*

□ Set current view as default

□ Select current record

Substance (hide)

| | |
|---|---|
| **ACD Registry Number** | 79670 |
| **MDL Number** | MFCD00190267 |
| **CAS Registry Number** | |
| **Chemical Name and Synonyms :**<br>- 5-AMINOLEVULINIC ACID-5-13C HYDROCHLORIDE<br>- 5-AMINOLEVULINIC-5-13C ACID HCL | |
| **Molecular Formula** | C5 H9 N O3 . Cl H |
| **Molecular Weight** | 168.58 |
| **Rule of Five** | 0 |
| **Computed partition coefficent (CLogP)** | -4.4 |
| **Molecular weight of largest fragment** | 132.119 |
| **Number of proton acceptors** | 4 |
| **Number of proton donors** | 2 |
| **Number of rotatable bonds (TDF)** | 3 |

Prices (hide)

ALDRICH (1 of 2)

| Catalog Number | Package Size | Package Price | Purity | Other Info |
|---|---|---|---|---|
| | | | | 5-AMINOLEVULINIC ACID-5-13C HYDROCHLORIDE |

Figure 3 shows the first part of a record retrieved in this search. Substance information, details about suppliers or their specific products (qualities, packages), and prices are displayed upon clicking the respective links; such a choice can be set as default during a search session with a check box. The left navigation frame shows all hitsets of the current session, as well as earlier and saved results. The „export records“ feature (see command line on top of display) permits downloading selected data fields (e.g., compound name, CAS RN, flag for bulk supplier, or new catalog entry) into an MS Excel spreadsheet, a feature we found very useful.

## 2.3 ChemACX

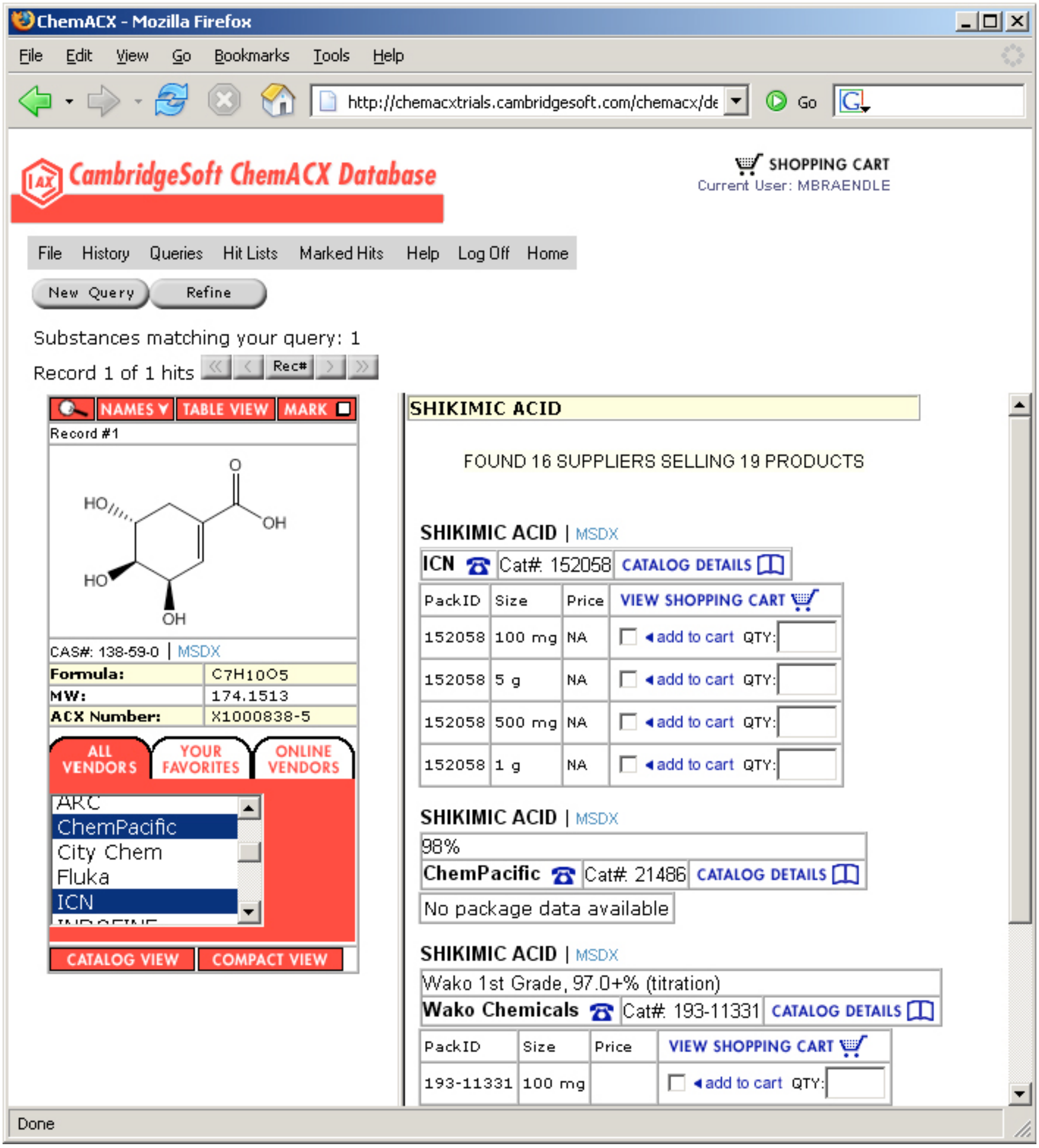
The Available Chemical Exchange database was created in 1998 by CambridgeSoft as a Web-based application. ChemACX 9.4 contains chemicals from 468 catalogs, with 983’718 product records representing 416'198 unique substances. For our tests, we had only access to version 9.2 with 352'253 compounds. The Internet version is updated twice a year, searched via a Java-enabled WWW browser and offers personal (one month or one year) as well as an one-year enterprise licenses. Up to 100 compounds may be exported into SD files from the Enterprise Internet Edition. In-house versions of the Enterprise Edition (with increased export facilities) are offered for MS Access and Oracle, and updated quarterly.

In an interesting licence model, ChemACX is also part of the ChemOffice Ultra software package on DVD for use with the desktop ChemFinder application provided by CambridgeSoft. This version additionally contains the screening compound directory ChemSCX and Material Safety Datasheets (ChemMSDX). Use is not restricted by a license period as for the Internet/Intranet versions, but there are no updates, and no export facilities. CambridgeSoft used to offer a free subset of the commercial database on the Web, but this is no longer available. Below the strucure (cf. Figure 4), a link „MSDX“ leads to Material Safety Data Sheets (MSDS) whenever this information is available.

For our evaluation we used the Enterprise Internet Edition. A compound like shikimic acid (cf. Figure 4 and Table 5) may be searched by structure in a specific „Exact“ mode, or in a more general „Full“ mode. With the correct configuration specified, both modes did not retrieve the desired compound, while a structure without „up“ and „down“ bonds in the „Full“ (but again not the „Exact“!) mode gave the record shown in Figure 4. Text searches are possible with CAS RN, molecular formula, molecular weight range, ACX or catalog number. All these facilities are presented to the user in a single and simple search mask.

From the scroll-down list of vendors in the display of search results (see Figure 4), up to five may be marked simultaneously, for display of supplier details, as shown for ChemPacific, ICN, and Wako in our example. A list of favorite suppliers can be defined to reduce the display which is extensive for commonly available compounds. Products may be collected in a shopping cart which is exportable as order forms in MS Excel format. The forms can then be completed and sent to the supplier.

**Figure 4.** Screenshot of Detailed Result View for shikimic acid in ChemACX Enterprise. Reproduced with permission of CambridgeSoft Corp.



ChemACX - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://chemacxtrials.cambridgesoft.com/chemacx/de
Go
CambridgeSoft ChemACX Database
SHOPPING CART
Current User: MBRAENDLE
File History Queries Hit Lists Marked Hits Help Log Off Home
New Query
Refine
Rec#
NAMES
TABLE VIEW
MARK
HO
O
OH
HO
OH
ALL VENDORS
YOUR FAVORITES
ONLINE VENDORS
ARC
ChemPacific
City Chem
Fluka
ICN
CATALOG VIEW
COMPACT VIEW
Done

Substances matching your query: 1

Record 1 of 1 hits

Record #1

CAS#: 138-59-0 | MSDX

| | |
|---|---|
| Formula: | C7H10O5 |
| MW: | 174.1513 |
| ACX Number: | X1000838-5 |

SHIKIMIC ACID

FOUND 16 SUPPLIERS SELLING 19 PRODUCTS

SHIKIMIC ACID | MSDX

ICN ☎ Cat#: 152058 CATALOG DETAILS

| PackID | Size | Price | VIEW SHOPPING CART |
|---|---|---|---|
| 152058 | 100 mg | NA | □ add to cart QTY: |
| 152058 | 5 g | NA | □ add to cart QTY: |
| 152058 | 500 mg | NA | □ add to cart QTY: |
| 152058 | 1 g | NA | □ add to cart QTY: |

SHIKIMIC ACID | MSDX

98%

ChemPacific ☎ Cat#: 21486 CATALOG DETAILS

No package data available

SHIKIMIC ACID | MSDX

Wako 1st Grade, 97.0+% (titration)

Wako Chemicals ☎ Cat#: 193-11331 CATALOG DETAILS

| PackID | Size | Price | VIEW SHOPPING CART |
|---|---|---|---|
| 193-11331 | 100 mg | | □ add to cart QTY: |

## 2.4 Chem Sources

Chem Sources as a printed „catalog of catalogs“ is the oldest of the sources discussed here. In electronic form, it became first available on the host STN International [6] which also offers the complete literature, structure, and reaction databases produced by the Chemical Abststracts Service, including the catalog database CHEMCATS (cf. Chapter 2.1). As an exception among catalog databases, the compound and producer information is held in two separate files, CSCHEM [7,8] and CSCORP [9,10], respectively. These are linked by the retrieval language STN Messenger and its cross-file search commands, permitting the transfer of supplier codes displayed with the desired compounds in CSCHEM to retrieve the full addresses in CSCORP. No price and package information is given for the products. CAS RNs were assigned to 132'802 of the total 203'303 product records in CSCHEM. This allows - as in the case of CHEMCATS (cf. Chapter 2.1) - to use all search features of the CAS Registry structure database (structure, substructure, molecular formula, molecular weight, elemental composition) in a preliminary compound search, limiting the compounds to those with records in CSCHEM, and then to transfer the CAS RNs retrieved into CSCHEM. Our test searches demonstrated (cf. Table 5), however, that this powerful and rather seamless method does by no means obviate the need for direct searches in CSCHEM via product name or name fragment, company name, or trade name classification. Direct searches with known CAS RNs are of course also possible. CSCHEM and CSCORP are produced by Chemical Sources International, and updated only once a year.

We used the STN version of Chem Sources for our comparison, but Chem Sources is also accessible via a WWW interface [11] that permits searches by chemical or trade name, molecular formula, CAS RN, supplier, or application, but not by structure. Searching in Chem Sources is possible via a guest account, but supplier information is only provided for subscribers. Print and CD-ROM editions are also available.

### 2.5 Chemicals Available for Purchase (CAP)

CAP [12] is a relative latecomer in the scene, obviously produced by Accelrys to round of its array of reaction and compound databases and chemical information management systems. It contains 568'712 unique compounds („reagents“) from 94 catalogs, as well as 3.1 million screening compounds from the same sources. Similar to ChemACX, CAP is offered in-house for Oracle, as is the case for other Accelrys reaction and compound databases. Because this database is accessible via local installations, but not via the web or a host, we did not include it in our evaluation.

**Table 1a.** Features of commercial chemicals directories (as of October 2006)

| Product | CHEMCATS | | ChemACX | ACD | ACD-SC | CSCHEM |
|---|---|---|---|---|---|---|
| Interface | SciFinder (Scholar) | STN | ChemACX Enterprise | DiscoveryGate | | STN |
| Producer | CAS | | CambridgeSoft | Elsvevier MDL | | Chemical Sources Int. |
| Category | Catalog DB | | Catalog DB | Catalog DB | DB of Screening Compounds | Catalog DB |
| **Content** | | | | | | |
| Number of products | 11,500,000 | | 983,000 | 1,400,000 | 6,000,000 | n.a. |
| Number of compounds | 3,400,000 | | 416,000 | 510,000 | 3,400,000 | 203,000 |
| Number of Suppliers | 828 | | 468 | 685 | 46 | 7,200 |
| Update frequency | Weekly | | 2/year | 4/year | 4/year | 1/year |
| Growth/Change per update (compounds) | n.a. | | ~13,000 | ~31,000 | ~207,000 | n.a. |
| Compound classes | Organics, Inorganics, Metallorganics, Polymers, Biopolymers, Materials | | Organics, Inorganics, Metallorganics, Polymers, Biopolymers, Materials | Organics, Inorganics Metallorganics, Polymers, Biopolymers, Materials | Organics, Metallorganics | Organics, Inorganics, Metallorganics, Polymers, Biopolymers, Materials |
| Compound Names | ✓ | | ✓ | ✓ | ✓ | ✓ |
| Structures | ✓ | | ✓ | ✓ | ✓ | ✓ |
| Supplier Adresses | ✓ | | ✓ | ✓ | ✓ | ✓ |
| Package Info | ✓ | | ✓ | ✓ | ✓ | ✗ |
| Product Prices | ✓ | | ✓ | ✓ | ✓ | ✗ |
| Purity | ✗ | | ✓ | ✓ | ✓ | ✗ |
| Physical data | ✗ | ✗ | ✗ | via DBs in DiscoveryGate | | ✗ |
| Safety (MSDS) | ✗ | ✗ | ✓ | via OHS MSDS DB in DiscoveryGate | | ✗ |
| Regulations | via CHEMLIST in SciFinder | ✗ | ✗ | ✗ | ✗ | ✗ |
| Further (Prep., Lit. Ref.) | via CA in SciFinder | ✗ | ✗ | via DBs in DiscoveryGate | | ✗ |
| **Access** | | | | | | |
| Part of a larger system | ✓<br>SciFinder | (✓)[1]<br>STN DBs | ✗ | ✓<br>MDL DiscoveryGate | | (✓)[1]<br>STN DBs |
| Database Medium | Host | Host | Web | Web | | Host |
| Price models | Ind./Acad. Subscription | Ind./Acad. Subscr. | Ind./Acad. Subscription | Ind./Acad. Subscription | | Ind./Acad. Subscr. |
| Licensing models | Task packages / fix fee | pay per use | fix fee | fix fee | | pay per use |
| **Technical requirements** | | | | | | |
| Platforms | Win/OS X | Win/OS X | Win/OS X | Win/OS X | | Win/OS X |
| Installation | ++ / –[2] | ++ / ++ | + / + | ++ / ++ | | ++/++ |
| Used Software, Plugins | SciFinder client | STN Express Client (Win) / Browser (STN Easy, STN on the Web ) | ChemDraw Plugin | IE / Safari Java | | Client (Win) |
| Browser compatibility | n.a. | n.a.[3] | Win: IE, Firefox<br>OS X: Safari | Win: IE<br>OS X: Safari | | n.a.[3] |

**Table 1a.** (continued)

| **Product** | **CHEMCATS** | | **ChemACX** | **ACD ACD-SC** | **CSCHEM** |
|---|---|---|---|---|---|
| Interface | SciFinder (Scholar) | STN | ChemACX Enterprise | DiscoveryGate | STN |
| **Search options** | | | | | |
| Chemical Names | (✓)[4] | ✓ | ✓ | ✓ | ✓ |
| Molecular Formula | (✓)[4] | (✓)[4] | ✓ | ✓ | (✓)[4] |
| Molecular Weight | (✓)[4] | (✓)[4] | ✓ | ✓ | (✓)[4] |
| CAS Registry Number | (✓)[4] | ✓ | ✓ | ✓ | ✓ |
| Suppliers | ✗ | ✓ | (✓)[5] | ✓ | ✓ |
| Prices | ✗ | ✗ | ✗ | ✗ | ✗ |
| Exact structure | (✓)[4] | (✓)[4] | ✓ | ✓ | (✓)[4] |
| Substructure | (✓)[4] | (✓)[4] | ✓ | ✓ | (✓)[4] |
| Similarity Search | (✓)[4] | ✗ | ✓ | ✓ | ✗ |
| Properties search | (✓)[4] | (✓)[4] | ✗ | ✓ | (✓)[4] |
| Bulk/fine flag | ✗ | ✗ | ✗ | ✗ | ✓ |
| Batch scripts | ✗ | ✓ | ✗ | ✗ | ✓ |
| Import | ✓[6] | ✓ | ✓[7] | ✓ | ✓ |
| - .mol files | ✓ | ✗ | ✓[7] | ✓ | ✗ |
| - Smiles | ✗ | ✗ | ✗ | ✗ | ✗ |
| **Special functions** | | | | | |
| External links | ✗ | ✗ | ✓ (MSDS) | ✓ | ✗ |
| Request for quote | ✗ | ✗ | ✗ | ✗ | ✗ |
| Direct ordering | ✗ | ✗ | ✗ | ✗ | ✗ |
| Indirect ordering | ✗ | ✗ | ✓ | ✗ | ✗ |
| Import/Export[6] | ✓ / ✓ | ✓ / ✓ | ✓ / ✓ | ✓ / ✓ | ✓ / ✓ |
| Integration into chemical information management systems (CIM) | ✗ | ✗ | ✓ | ✗[8] | ✗ |
| Current awareness (SDI) | SciFinder, for substances and references | ✗ | ✗ | (✗)<br>New Compounds can be searched with flag NEW | ✗ |
| **User Interface** | | | | | |
| Search mask | + | + | + | ++ | + |
| Structure editor | ++ | + | + | ++ | + |
| Display of hits | + | + | ++ | ++ | + |
| Navigation | – | n.a. | ++ | ++ | n.a. |
| History | ✗ | ✓ | ✓ | ✓ | ✓ |
| Sorting | ✗ | (✓) | ✗ | ✓ | (✓) |
| **User support** | | | | | |
| Documentation | ++ | ++ | + | + | ++ |

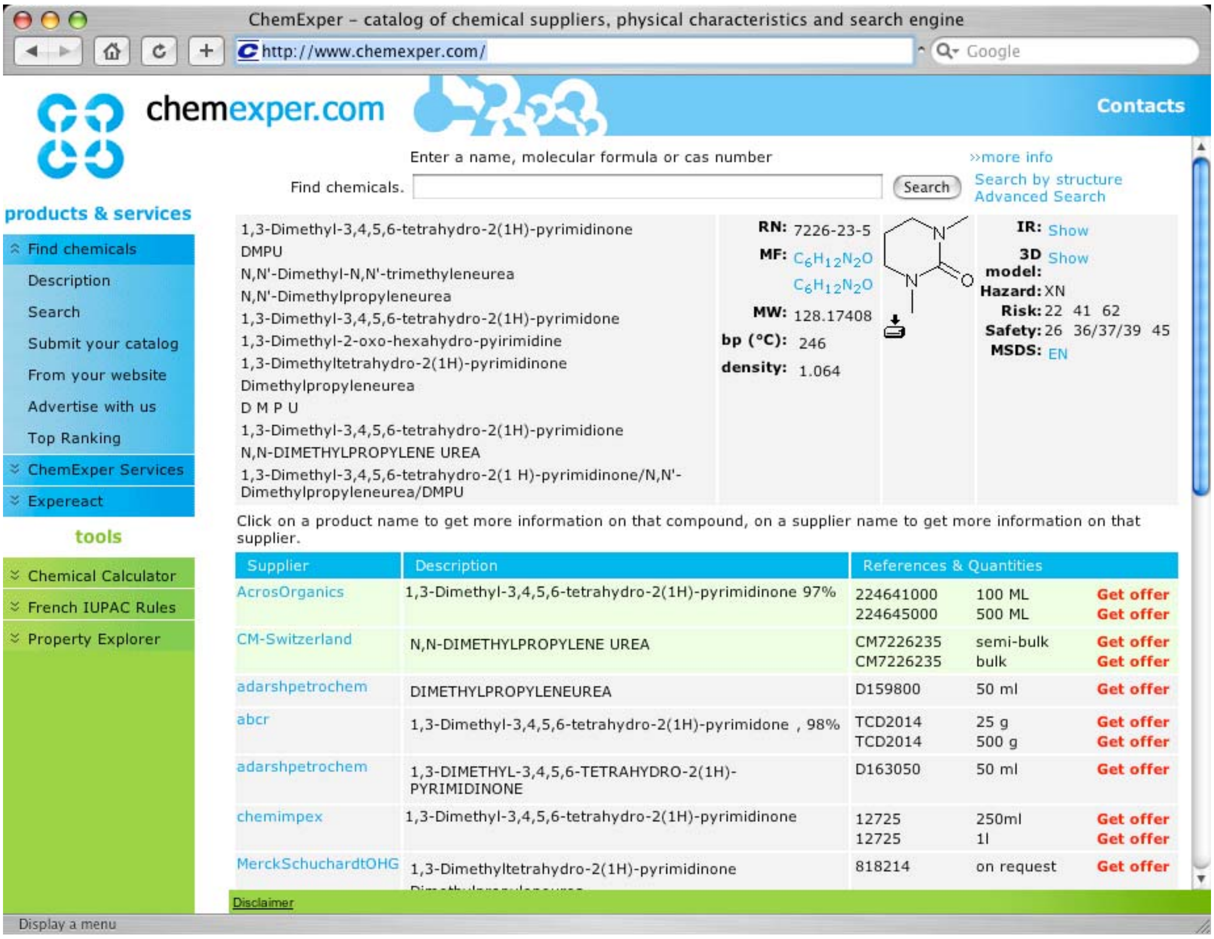
n.a = information not available, ✓ = exists, (✓) = partially exists, ✗ = does not exist/can not be done, ++ = very good, + = good , – = unsatisfactory, – – = unacceptable

[1] can be used as only database, but also together with other hosted databases
[2] see chapter 2.1
[3] STN for the Web: information n.a. for current browsers
[4] via CAS Registry only
[5] by way of a list of favorite vendors
[6] max. 25 names or CAS RNs via Locate dialog, or 1 structure via structure editor
[7] via ChemDraw structure ditor
[8] ACD/ISIS can be integrated

## 2.5 ChemExper

Of the free chemicals directories available on the WWW, ChemExper [13] produced by the homonymous company ChemExper sprl probably comes closest to the purpose and functionality of the major commercial sources. This rapidly growing database lists about 200,000 unique compounds, 819,000 product records, and more than 1000 suppliers. In addition, about 10,000 MSDS, over 10,000 IR and NMR spectra, and about 7000 3D models that can be visualized in a JMOL applet [14] are provided, a unique feature among the sources examined. Updates occur on a daily basis as we found in test searches. The business model of ChemExper relies on Google Adwords [15] and the option for suppliers to bid for better ranking in the display of  product details.

**Figure 5.** Screenshot of Detailed Results Display of ChemExper for DMPU. Reproduced with permission of ChemExper sprl



Enter a name, molecular formula or cas number

Find chemicals.

1,3-Dimethyl-3,4,5,6-tetrahydro-2(1H)-pyrimidinone
DMPU
N,N'-Dimethyl-N,N'-trimethyleneurea
N,N'-Dimethylpropyleneurea
1,3-Dimethyl-3,4,5,6-tetrahydro-2(1H)-pyrimidone
1,3-Dimethyl-2-oxo-hexahydro-pyirimidine
1,3-Dimethyltetrahydro-2(1H)-pyrimidinone
Dimethylpropyleneurea
D M P U
1,3-Dimethyl-3,4,5,6-tetrahydro-2(1H)-pyrimidione
N,N-DIMETHYLPROPYLENE UREA
1,3-Dimethyl-3,4,5,6-tetrahydro-2(1 H)-pyrimidinone/N,N'-Dimethylpropyleneurea/DMPU

**RN:** 7226-23-5
**MF:** $C_6H_{12}N_2O$ $C_6H_{12}N_2O$
**MW:** 128.17408
**bp (°C):** 246
**density:** 1.064

**IR:** Show
**3D model:** Show
**Hazard:** XN
**Risk:** 22 41 62
**Safety:** 26 36/37/39 45
**MSDS:** EN

Click on a product name to get more information on that compound, on a supplier name to get more information on that supplier.

| Supplier | Description | References & Quantities | | |
|---|---|---|---|---|
| AcrosOrganics | 1,3-Dimethyl-3,4,5,6-tetrahydro-2(1H)-pyrimidinone 97% | 224641000<br>224645000 | 100 ML<br>500 ML | Get offer<br>Get offer |
| CM-Switzerland | N,N-DIMETHYLPROPYLENE UREA | CM7226235<br>CM7226235 | semi-bulk<br>bulk | Get offer<br>Get offer |
| adarshpetrochem | DIMETHYLPROPYLENEUREA | D159800 | 50 ml | Get offer |
| abcr | 1,3-Dimethyl-3,4,5,6-tetrahydro-2(1H)-pyrimidone , 98% | TCD2014<br>TCD2014 | 25 g<br>500 g | Get offer<br>Get offer |
| adarshpetrochem | 1,3-DIMETHYL-3,4,5,6-TETRAHYDRO-2(1H)-PYRIMIDINONE | D163050 | 50 ml | Get offer |
| chemimpex | 1,3-Dimethyl-3,4,5,6-tetrahydro-2(1H)-pyrimidinone | 12725<br>12725 | 250ml<br>1l | Get offer<br>Get offer |
| MerckSchuchardtOHG | 1,3-Dimethyltetrahydro-2(1H)-pyrimidinone | 818214 | on request | Get offer |

The Google-like "Quick Search" bar accepts chemical names, CAS RNs and molecular formulas. Abbreviations can be used in molecula formulas for common functional groups and substituents. Structure and substructure searches can be carried out in the "Advanced Search" window, as well as searches for physical properties (bp, mp, density, IR, NMR), suppliers, creation and modification dates. For structure input, the JME Molecular Editor by Peter Ertl from Novartis Pharma [16] is used. Upon search, a table of compounds with chemical names, CAS RNs, and structures is returned. Results can be sorted by clicking either on the “IUPAC name” or on the “RN (CAS)” column headers. By clicking the harddisk icon, molfiles can be downloaded. Clicking a result row leads the user to the detailed display of the compound, its properties, and the list of product records. Figure 5 shows such a typical compound record, with links to an IR spectrum and a MSDS. The product table can be sorted either by product description or reference. Product prices are not given, but can be requested by the “Get offer” link from the suppliers that opens a form sent by way of the ChemExper webserver directly to the supplier.

### 2.6 ChemNet

ChemNet is a buyers and sellers portal for chemical and pharmaceutical products [17]. Founded 1995 in the US, it changed owners a few times and since 2001 belongs to the chinese enterprise Zheijang NetSun Co. Ltd. According to the producer, it encompasses more than 300,000 products and 80,000 suppliers worldwide. The homepage, cluttered with disturbing blinking ads, offers three search modes for products, websites, and companies. The search engine is text-based and therefore allows to find only chemical and trade names, CAS RNs, molecular formula and suppliers. No indication is given on the website, however, that these data types can be searched. A list of retrieved products is first shown, from which the desired product can be selected to bring up a page of suppliers. There, links to request a quote from a supplier and links to the supplier web sites are offered, as well as the handy possibility to filter vendors by region.

### 2.7 eMolecules

eMolecules, formerly known as Chmoogle, was launched in November 2005 by eMolecules Inc. [18] This search engine for chemical structures gathers its information by both spidering the WWW for chemical information (16.4 million sources) and including catalog data submitted from about 100 suppliers. It contains over 5.6 million unique compounds, 500,000 CAS RNs, and 2 million IUPAC, common and trade names. The exact number of commercially available compounds was not available. Three search modes, Basic, Advanced and Expert Search are offered. Basic Search permits searches on exact and partial structure, chemical and trade names, and CAS RN. Advanced Search in addition allows to limit by supplier and commercial availability. For structure searching, again the JME Molecular Editor [16] is provided as default, but ChemDraw [19] and ISIS/Draw [20] can be chosen in the preferences as alternative structure editors. Use of the latter is limited to Internet Explorer and the MS Windows operating system, and our tests showed that transfer of the structures drawn in these editors to the search bar did work without problems only in Expert mode. Searching in the Basic and Advanced search modes is very fast, irrespective of exact or substructure retrieval. A list of structures together with product numbers and supplier information is displayed in the results. No price and package information is given. Ranking of compounds found in substructure searches is unclear, and no information about ranking was available in the help. The supplier links open a separate window with supplier information. If possible, deep links via product number to compound information of a source or supplier are provided. The "edit" link loads the displayed structure in the structure editor for further searches. Similarly, the "details" link below the structure image opens a new window with SMILES, chemical name and CAS RN information. "G" and "Y!" links are provided for the latter two to continue with text searches in Google and Yahoo.

Expert Search, which needs previous user registration, seems to be a true Web 2.0 application, since several searches can be carried out at once in the background and one is still able to continue work on the history of searches and on the hitlists. Search options are freely configurable, as well as the columns to be displayed in the hitlists. Hitlists can be combined with Boolean logic, compounds can be cherry-picked, and results may be exported in SMILES format, or as MDL SD file or MS Excel spreadsheets for which data fields to be exported can be chosen. Batch operations are possible by recording, storing and replaying macros.

## 2.8 ICISsearch

As ChemNet, ICISsearch [21] is a portal for chemical products and companies produced by Reed Business Information, a member of Reed Elsevier. The search engine is provided by the global supplier search facility KellySearch. The web pages looks tidy, however, many links ("List My Company", "Advertising Solutions", "About Us", "Contact Us" etc.) were found to be faulty. Searching is only text based and allows to find product names, services, CAS RNs or companies. Companies can be filtered by country or region. As with Google, sponsored links are displayed in a separate box in the list of results. Clicking a product link results in a list of suppliers which can be filtered by region or country if desired, and on which links for company information and request for quotes are offered. No product price information is given. The company name link either opens an address and product information page provided by ICISsearch, or loads the product information pages of the respective company in a frame and presents it under the URL of ICISsearch. The latter also occurs with the "more information” link. Presenting external information in frames is considered bad practice in web design and may even come into conflict with business regulations of certain countries (e.g., Germany).

## 2.9 LabVelocity

The LabVelocity portal produced by LabVelocity Inc., an affiliate of Infotrieve Inc., allows to find products, protocols, literature and news in life sciences [22]. Products are ordered in categories; the chemicals category offers search for 22,089 chemicals by name, CAS RN, company, molecular weight and molecular formula, the last, however, not being very useful since nearly no molecular formulas were assigned to products. Price information is displayed when one is logged in as a registered user (registration is free). If available, the currency displayed is taken from the country set in the user’s preferences. Products can be added to an order list which is used to generate order forms. Among the chemicals catalogs discussed in this article, LabVelocity is the only one which allows sorting products by price or comparing prices of cherry-picked records. A request information function to get quotes complements the very user-friendly and tidy interface.

## 2.10 Super Natural Database

Super Natural Database [23] produced by the Charité at the Medical Faculty of the Humboldt University in Berlin contains 45,917 purchasable natural compounds (8 suppliers) with 2.665.881 conformers generated with Accelrys Catalyst [24], and seems to be intended for drug screening. Compounds are searchable by chemical name, CAS RN, molecular formula, molecular weight, log P, template search via template substructure, supplier name, and structure similarity search. Structures can be entered with the commercial ChemAxon MarvinSketch Java applet [25], a function-rich structure editor possessing a multitude of structure templates. For display of results, the MarvinView Java applet [25] is provided as an option to select; otherwise MDL Chime [26] must be installed, which restricts use to MS Windows and Internet Explorer only. Results are ranked by similarity; product ID and supplier name are given, but no link the respective supplier. One of the major drawbacks of this database is that no exact and substructure search options exist. Our tests showed that its similarity retrieval searches too broadly and yields far too many useless structures that don’t have any topological features in common with the structure entered. Further, display of the number of retrieved records, record navigation, and documentation about the similarity search method are missing.

## 2.11 ZINC

ZINC [27,28], provided by the Shoichet Laboratory in the Department of Pharmaceutical Chemistry at the University of California, San Francisco, is a free database containing 4.3 million compounds for virtual screening, of which 2.67 million are commercially available from 45 suppliers. Compounds are retrieved by substructure, vendor, purchasibility, availability, molecular weight, various physical properties such as log P, but not by chemical name or CAS RN. The JME Molecular Editor [16] is used for structure input. Alternatively, a list of SMILES [29] strings or ZINC codes can be uploaded for batch search. Substructure search is the default; it can be limited approximately to exact search by specifying a molecular weight range or adding a Tanimoto similarity index of 100 to the SMILES string. The result list displays the found structures, vendor names and product ID, and physical properties. Useful deep links to product records on the vendor's websites are offered. Clicking a structure image opens a window for 3D visualization of the compound. Records can be navigated forward and backward, but no total number of found records is given. Sets of records can be downloaded in SMILES, molfile, SD file, and DOCK flexibase formats. Also available for download are predefined subsets of compounds on a separate webpage.

## 2.12 Various

There are many other chemical product catalogs in the internet including the individual supplier's webshops that we were not able to discuss for the number alone. For example, we left out NIH's PubChem [30], one of the largest compound database on the WWW, because it does not contain direct information to suppliers, but to suppliers only via other catalog databases (e.g. ChemExper). To get an overview of product catalogs on the WWW, we recommend to use Web directories such as the directories of Google [31], DMOZ Open Directory [32], or directories offered by industrial associations, e.g. the directory by the German Verband der Chemischen Industrie, Regionalverband Nordrhein-Westfalen [33].

**Table 1b.** Features of free chemicals directories (as of October 2006)

| Product | ChemExper | ChemNet | eMolecules | ICISsearch | Lab-Velocity | Super Natural DB | ZINC |
|---|---|---|---|---|---|---|---|
| Producer | ChemExper sprl | Zhejiang NetSun Co., Ltd. | eMolecules Inc. | Int. Chemical Information Services, Reed Elsevier | Infotrieve Inc. | Charité, Medical Faculty, Humboldt University Berlin | Univ. California, San Francisco |
| Category | Catalog DB | Portal | Search Engine | Portal | Portal | DB of Natural Compounds | DB of Screening Compounds |
| **Content** | | | | | | | |
| Number of products | 819,000 | n.a. | n.a. | n.a. | 22,000[1] | n.a. | 8,573,000 |
| Number of compounds | 200,000 | n.a. | 5,600,000 | n.a. | n.a. | 45,917 | 4,300,000 (2,667,000)[2] |
| Number of Suppliers | 1078 | n.a. | 102 | n.a. | n.a. | 8 | 45 |
| Update frequency | daily | n.a. | n.a. | n.a. | n.a | 2/year | n.a.[3] |
| Growth/Change per update (compounds) | ~20,000/ months | n.a. | n.a. | n.a. | n.a | n.a | n.a. |
| Compound classes | Organics, Inorganics, Metallorganics, Polymers, Biopolymers, Materials | Organics, Inorganics, Metallorganics, Polymers, Biopolymers, Materials | Organics | Organics, Inorganics, Metallorganics, Polymers, Biopolymers, Materials | Organics, Inorganics, Metallorganics, Polymers, Biopolymers, Materials | Organics | Organics, Metallorganics |
| Compound Names | ✓ | ✓ | ✓ | ✓ | ✓ | (✓) | ✗ |
| Structures | ✓ | ✗ | ✓ | ✗ | ✗ | ✓ | ✓ |
| Supplier Adresses | ✓ | ✓ | ✓ | ✓ | ✓ | (✓) | ✓ |
| Package Info | ✓ | ✗ | ✗ | ✗ | ✓ | ✗ | ✗ |
| Product Prices | ✗ | ✗ | ✗ | ✗ | ✓[4] | ✗ | ✗ |
| Purity | ✓ | ✗ | ✗ | ✗ | ✓ | ✗ | ✗ |
| Physical data | ✓ | ✗ | ✗ | ✗ | ✗ | ✓ | ✓ |
| Safety (MSDS) | ✓ | ✗ | ✗ | ✗ | ✓ | ✗ | ✗ |
| Regulations | ✓ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ |
| Further (Prep., Lit. Ref.) | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ |
| **Techn. requirements** | | | | | | | |
| Used Software, Plugins | none | none | none | none | none | MDL Chime (IE Win), Marvin Java applets | none |
| Browser compatibility | IE, Firefox (Win) Safari, Firefox (OS X) | IE, Firefox (Win) Safari, Firefox (OS X) | IE, FireFox (Win) Safari, Firefox (OS X) | IE, Firefox (Win) Safari, Firefox (OS X) | IE, Firefox (Win) Safari, Firefox (OS X) | IE, Firefox (Win) Safari, Firefox (OS X) | IE, Firefox (Win) Safari, Firefox (OS X) |
| Structure Editor | JME Molecular Editor | n.a. | JME Molecular Editor; ChemDraw, ISISDraw (Win IE only) | n.a. | n.a. | ChemAxon Marvin-Sketch applet | JME Molecular Editor[5] |

**Table 1b.** (continued)

| **Product** | **ChemExper** | **ChemNet** | **eMolecules** | **ICISsearch** | **Lab-Velocity** | **Super Natural DB** | **ZINC** |
|---|---|---|---|---|---|---|---|
| **Search options** | | | | | | | |
| Chemical Names | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✕ |
| Molecular Formula | ✓ | ✓ | ✓[6] | ✕ | (✓)[7] | ✓ | ✕ |
| Molecular Weight | ✓ | ✕ | ✓[6] | ✕ | (✓)[7] | ✓ | ✓ |
| CAS Registry Number | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✕ |
| Suppliers | ✓ | ✓ | ✓[6] | ✓ | ✓ | ✓ | ✓ |
| Prices | ✕ | ✕ | ✕ | ✕ | ✕ | ✕ | ✕ |
| Exact structure | ✓ | ✕ | ✓ | ✕ | ✕ | ✕ | (✕)[8] |
| Substructure | ✓ | ✕ | ✓ | ✕ | ✕ | ✕ | ✕ |
| Similarity Search | ✕ | ✕ | ✕ | ✕ | ✕ | ✓ | ✓ |
| Properties search | ✓ | ✕ | ✕ | ✕ | ✕ | ✓ (log P) | ✕ |
| Bulk/fine flag | ✕ | ✕ | ✕ | ✕ | ✕ | ✕ | ✕ |
| Batch scripts | ✕ | ✕ | ✓[6] | ✕ | ✕ | ✕ | ✕ |
| Import | ✕ | ✕ | ✓[6] | ✕ | ✕ | ✓ | ✓ |
| - .mol files | ✕ | ✕ | ✓[9] | ✕ | ✕ | ✓[10] | ✕ |
| - Smiles | ✕ | ✕ | ✕ | ✕ | ✕ | ✓[10] | ✓ |
| **Special functions** | | | | | | | |
| External links | ✓ | ✓ | ✓ | ✓ | ✓ | ✕ | ✓ |
| Request for quote | ✓ | ✓ | (✓) | ✓ | ✓ | ✕ | ✓ |
| Direct ordering | ✕ | ✕ | ✕ | ✕ | ✕ | ✕ | ✕ |
| Indirect ordering | ✕ | ✕ | (✓)[11] | ✕ | ✓ | ✕ | ✕ |
| Import/Export | ✕ | ✕ | ✕ (✓)[6] | ✕ | ✕ | ✕ | ✓ |
| Integration into CIM systems | ✕ | ✕ | ✕ | ✕ | ✕ | ✕ | ✕ |
| Current awareness SDI) | ✕ | ✕ | ✕ | ✕ | ✕ | ✕ | ✕ |
| **User Interface** | | | | | | | |
| Search mask | ++ | – | ++ | + | + | + | + |
| Structure editor | + | ✕ | + | ✕ | ✕ | + | + |
| Display of hits | ++ | – | ++ | + | + | – | + |
| Navigation | – | – – | + | – – | – | – – | – – |
| History | ✕ | ✕ | (✓)[6] | ✕ | ✕[12] | ✕ | ✕ |
| Sorting | ✓ | ✕ | (✓)[6] | ✕ | ✓ | ✕ | ✕ |
| **User support** | | | | | | | |
| Documentation | – | ✕ | ++ | ✕ | ++ | + | – |

n.a = information not available, ✓ = exists, (✓) = partially exists, ✕ = does not exist/can not be done, ++ = very good, + = good , – = unsatisfactory, – – = unacceptable

[1] Chemicals category only
[2] Purchasable compounds
[3] Exported subsets: 4/year
[4] after registration and login
[5] Input of „nonstandard“ atoms did not work
[6] in Expert Mode
[7] searchable, bot not available in majority of chemicals
[8] approximatively via Tanimoto threshold
[9] via ChemDraw or ISIS/Draw
[10] via structure editor
[11] via deep links to vendor
[12] History for order forms exists

## 3 Searches and Results

### 3.1 Selected compounds

To accommodate the wide variety of product categories in chemicals catalogs, we selected typical example compounds for our benchmark searches. As categories,  bulk chemicals, natural products, pharmaceuticals, isotope-labelled compounds, enzymes, polymers, reagents, inorganic materials, and compound classes were considered. These categories allow to test the different search options the databases offer. Table 2 lists the searched compounds and their uses. Propylene glycol is a bulk chemical widely used as antifreeze, deicer, brake fluid, polymer additive and in pharmaceuticals and cosmetics [34]. Here, we looked for 1,2-propanediol, its stereoisomers and labelled derivatives. For L(-)-glucose, the enantiomer of the common glucose, we searched both the pyranose and aldose forms.  For δ-aminolevulinic acid, we were interested in isotope-labelled compounds only, as used in metabolism studies. Enzymes such as pig liver esterase or phospholipase A2 were selected due their importance in biotechnology and because of lack of structure they are difficult to search with regards to name variations. Taddols and chiral phosphines were chosen to test the databases for their ability to perform substructure searches including stereochemistry. Chloromethyl polystyrene, also known as Merrifield’s peptide resin, used in automated synthesis of peptides, oligosaccharides and oligonucleotides, and PEG 8000 served as examples for rather tricky polymer searches.  Solid state compounds such as the hard material TiN, the superconducting La- Ba-Cu oxides, or zeolites utilized as molecular sieves or as industrial catalysts in petroleum cracking, are good benchmark compounds for element composition searches.

**Table 2.** Compounds used in searches

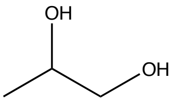
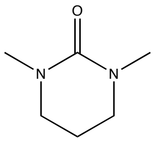
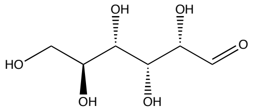
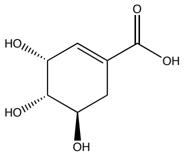
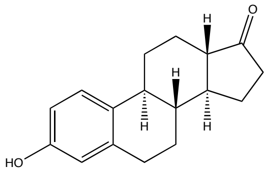
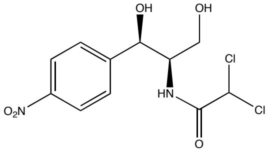
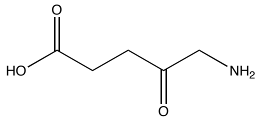
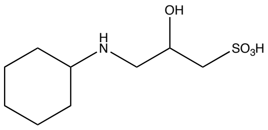
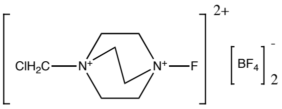
| Name | Structure | Type | Usage | CAS RN | Query |
|---|---|---|---|---|---|
| Propylene glycol |  | bulk chemical, solvent | Automobile industry, polymer additive, pharma-ceuticals, cosmetics[34] | 57-55-6 | Structure |
| N,N'-Dimethyl-propyleneurea DMPU |  | bulk chemical, solvent | Solvent | 7226-23-5 | Name<br>Structure |
| L-(-)-glucose |  | Natural product | Sweetener | 921-60-8<br>1992-85-5 | Name<br>Structure |
| Shikimic acid |  | Natural product, bulk | Pharmaceutical intermediate | 138-59-0 | Name<br>Structure |
| Estrone |  | Hormone, pharmaceutical | Pharmaceutical | 53-16-7 | Structure |
| Chloramphenicol |  | Pharmaceutical | Antibiotics | 56-75-7 | Name<br>Structure |
| δ-aminolevulinic acid |  | Isotope labelled compound | Biosynthetic intermediate, Metabolism studies | 106-60-5 | Name<br>Structure |
| Pig liver esterase | | Enzyme | Stereoselective synthesis | 9016-18-6 | Name |
| Phospholipase A2 | | Enzyme | Food chemistry, Biotechnology | 9001-84-7 | Name |
| Chloromethyl polystyrene | | Polymer | Polymer | 29296-32-0<br>70024-51-0<br>121961-20-4 | Name<br>Structure<br>Molecular formula |
| Polyethylene-glycol 8000 | | Polymer | Polymer | | Name |
| CAPSO |  | Buffer | Buffers | 73463-39-5 | Name<br>Structure |
| Zinc borohydride | $Zn(BH_4)_2$ | Reagent | Chemoselective reductions | 17611-70-0 | Name<br>Structure |
| Selectfluor |  | Reagent | Pharmaceutical applications | 140681-55-6<br>140681-68-1 | Name<br>Structure |
| $Yb(OTf)_3$ | $Yb(OTf)_3$ | Reagent | Stereoselective synthesis | 54761-04-5 | Structure |
| TiN | | Inorganic solid | Coatings | 25583-20-4 | Element composition |

**Table 2.** (continued)

| Name | Structure | Type | Usage | CAS RN | Query |
|---|---|---|---|---|---|
| Zeolites | | Inorganic solid | Catalysis, molecular sieves | | Name |
| Taddols | | Chiral reagents | Stereoselective synthesis | | Substructure |
| Chiral phosphines | | Chiral reagents | Stereoselective synthesis | | Substructure |
| Tertiary alkyliodides | | Compound class | | | Substructure |
| La-Ba-Cu oxides | | Compound class | High-Tc Superconductivity | | Element composition |
| Disubstituted ferrocenes | | Compound class, metallorganic compound | | | Substructure |

### 3.2 Results (Cross comparison for DMPU and phospholipase A2 in all databases)

Due to cost and time restrictions, it was not possible for us to carry out the searches for all substances in all conceivable modes (structure, CAS RN, name) in all databases. However, we did searches in all databases for DMPU and phospholipase A2. Tables 3a and 3b compare numbers of found product records and suppliers for commercial and free databases, respectively. Comparison of product record numbers is hampered by the fact that the definition of product record is different for each database. E.g., in CHEMCATS, a product record may correspond to single package size of a product, it may also correspond to one particular product spanning several different package sizes and prizes. In ACD (MDL DiscoveryGate), product records are differentiated by different catalog numbers. Hence, numbers must be taken with a grain of salt. For DMPU, CHEMCATS yields the most products and suppliers, followed by ACD and ChemACX. The web-based ChemExper provides the most data on products and suppliers; however, prices are not listed and must be requested from the supplier. The acronym retrieved the desired solvent in CHEMCATS (STN and SciFinder Scholar) and ACD (MDL DiscoveryGate, ChemACX), but not in STN CSCHEM, while the CAS RN for this solvent worked in most of the databases examined. In STN, searching for “DMPU” in CAS Registry and then via the retrieved CAS RN in CHEMCATS yielded 20 suppliers, while “DMPU” in STN CHEMCATS itself gave only two with this name search. This is one of the examples where CAS RNs are more efficient than names, an observation typical in our experience for “straightforward” organic compounds. As other examples show (see below), polymers, materials and inorganics are quite different in this respect, and even for organics a supplementary name search may be in order – but only as a supplement, not as a replacement for a structure/CAS RN search.

For phospholipase A2, where only name or CAS RN search is possible, the free databases only gave one or two products, if at all, whereas the commercial databases yielded several product records. IN STN CHEMCATS, a name search retrieved five suppliers for Phospholipases A2 from several species with no CAS RN, in addition to the 11 product records found with the CAS RN 9001-84-7 for Phospholipase A2. Of the seven phospholipase A2 records in STN CSCHEM, six had the CAS RN 9001-84-7 assigned, but a phospholipase A2 from Crotalus D. terificus carried no RN, and was thus only retrievable via a name search directly in STN CSCHEM, not via the standard search in CAS Registry, followed by RN crossover. We noticed several cases where enzymes had structures and CAS RNs of low-molecular weight compounds incorrectly assigned to them.

**Table 3a.** Comparison of search results for DMPU and Phospholipase A2 in commercial chemicals catalog databases. Numbers given as number of product records / number of suppliers (for definition see text)

| Substance | CHEMCATS | | ChemACX | ACD | ACD-SD | CSCHEM |
|---|---|---|---|---|---|---|
| | SFS | STN | | DiscoveryGate | | |
| DMPU | | | | | | |
| *Structure* | 23 / 20 | n.a. | 15 / 12 | 14[1] / 17 | 2 / 1 | n.a. |
| *CAS RN* | 23 / 20 | 23 / 20 | 15 / 12 | 14[1] / 17 | 0 / 0 | 3 / 10 |
| *Name* | 23 / 20 | 2 / 2 | 15 / 12 | 14[1] / 17 | 0 / 0 | 0 / 0 |
| Phospholipase A2 | | | | | | |
| *CAS RN* | 11 / 6 | 11 / 8 | 17 / 6 | 25 / 10 | 0 / 0 | 7[3] / 8 |
| *Name* | 11 / 6 | 30 / 10 | 17 / 6 | 27 / 10 | 0 / 0 | 7[3] / 7 |

[1] Number of products with different packaging: 33

[2] Number of products with different packaging: 17

[3] One record each retrieved exlusively with CAS RN, or name, respectively

**Table 3b.** Comparison of search results for DMPU and Phospholipase A2 in free chemicals catalog databases. Numbers given as number of product records / number of suppliers (for definition see text)

| Substance | ChemExper | ChemNet | eMolecules[1] | ICISsearch | Lab-Velocity | Super Natural | ZINC[1] |
|---|---|---|---|---|---|---|---|
| DMPU | | | | | | | |
| *Structure* | 26[2] / 19 | n.a. | 11 / 8 | n.a. | n.a. | 0 / 0 | 7 / 4 |
| *CAS RN* | 26[2] / 19 | 1 / 1 | 11 / 8 | 1 / 1 | 0 / 0 | 0 / 0 | n.a. |
| *Name* | 26[2] / 19 | 1 / 1 | - | 1 / 1 | 0 / 0 | 0 / 0 | n.a. |
| Phospholipase A2 | | | | | | | |
| *CAS RN* | 1 / 1 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | n.a. | n.a. |
| *Name* | 1 / 1 | 1 / 1 | 0 / 0 | 0 / 0 | 2 / 2 | n.a. | n.a. |

[1] PubChem links not counted

[2] Number of products with different packaging: 39

## 3.3 Further examples

Table 4 shows results for structure searches in the commercial catalog databases and ChemExper for propylene glycol. The number of compounds found by searching the partial structure clearly reflects the total size of the databases. With respect to exact structure searches, ChemACX matches well with SciFinder Scholar. For CAS RN searches, all databases except ChemExper find the 1,2-propanediol without stereo information, the (S)-(+) and the (R)-(-) stereoisomers, the perdeuterated, the deuterium-6 and the OD-labelled isomers. ACD retrieves two carbon labelled compounds as well. A polymeric glycol was retrieved only with ChemACX. CHEMCATS and ACD proved to yield the most product records and supplier data for the principal compound and its stereoisomers. This polymer is alos present in CHEMCATS (96 product records,19 suppliers), but registered as a SRU (Structurally Repeating Unit) and therefore not retrievable with the propylene glycol structure. In ChemExper, the RN for the (R) and (S) enantiomer retrieved the same, mixed result set; besides both enantiomers, it included two compound records marked „(R)“ or „(S)“ and two with no stereodescriptor.

**Table 4.** Comparison of structure search results for propylene glycol

| Type of structure search | | Number of compounds | | | | |
|---|---|---|---|---|---|---|
| | | SFS | ChemACX | ACD | CSCHEM[1] | ChemExper |
| | Substructure | 51,056[2] | 10,856 | 17,902[3] | 5,851 | 27,017 |
| | Exact | 7 | 2 | 1 | | 1 |
| | Full | | 7 | | | |
| | Similarity | 127 | | 4 | | |
| Compound | | Number of product records / suppliers | | | | |
| | | SFS | ChemACX | ACD | CSCHEM | ChemExper |
| 57-55-6 | | 112 / 51 | 68 / 37 | 103 / 61 | 4 / 102 | 61 / 33 |
| 4254-15-3 | (S)-(+) | 34 / 30 | 23 / 20 | 30 / 26 | 4 / 15 | 27 / 21[4] |
| 4254-14-2 | (R)-(-) | 31 / 30 | 17 / 17 | 26 / 14 | 2 / 15 | 24 / 22[4] |
| 52910-80-2 | d6 | 2 / 2 | 3 / 3 | 5 / 5 | 1 / 3 | 0 |
| 80156-55-4 | d8 | 3 / 3 | 2 / 2 | 5 / 5 | 1 / 4 | 0 |
| 58161-11-8 | $(\text{-OD})_2$ | 4 / 4 | 1 / 1 | 4 / 4 | 1 / 1 | 0 |
| 212575-35-4 | 1-$^{14}$C | 1 / 1 | 0 | 3 / 3[5] | 0 | 0 |
| | 1,2-(-$^{13}$C)$_2$ | 0 | 0 | 1 / 1 | | |
| 25322-69-4 | polymer | -[4] | 46 / 10 | | | |

[1] via CAS Registry

[2] only acyclic fragment, AutoFix option had to be used

[3] ACD-SD: 41,906

[4] see text

[5] no CAS RN

For L(-)-glucose, searches were carried out in SciFinder Scholar (26 product records with structure search), ACD (26 product records), and CSCHEM (4 product records). Here, SciFinder Scholar demonstrated its power in structure searching by retrieving automatically both the aldose and pyranose forms independent of the entered structure. Shikimic acid did not present problems, with one exception: While a name search in ChemACX retrieved both the natural stereoisomer of shikimic acid and 3-dehydroshikimic acid, a full structure search (the most comprehensive structure search option) was successful only if the structure query was entered without stereochemistry. A search for suppliers of estrone in STN CSCHEM via CAS RN or the name “estrone” gave the same result (4 catalog records). An additional search with the name “oestrone” retrieved a carbon-13 labelled estrone with no CAS RN assigned; this would thus be missed in a search for labelled estrones that usually start in CAS Registry.

Trying to retrieve isotope-labelled δ-aminolevulinic acids (δ-ALA) taught us some lessons in several catalog databases. In SciFinder Scholar, limiting the structure to “single component” may be dangerous, as, in this example, the compounds are mostly sold as hydrochlorides, which is common with amino acids and other amino compounds. Two commercially available labelled hydrochlorides were found. In ACD (DiscoveryGate), the choice of the structure search mode was shown to be critical to retrieve the desired labelled δ-ALA. Both the hydrochloride structure and the free acid had to be searched separately. For the hydrochloride, “Exact” search mode or “Include Isomers” gave two records, “Include Salts” additionally the free aminolevulinic acid, all unlabelled. Only “Include Tautomers” retrieved eight labelled hydrochlorides. „Include Tautomers“ search mode had also to be used to find the tritiated free acid. In the help, there was no indication that „Include Tautomers“ can be used to find isotope-labelled compounds. A name search for all compounds containing the name fragment “aminolevulinic” retrieved seven of the previous eight labelled ALA hydrochloride, the tritium-labelled free acid, as well as the unlabelled acid, its hydrochloride and some substituted derivatives. But this in turn missed the 5-AMINO-15N-LEVULINIC ACID HCL retrieved via the structure with “Include Tautomers”. In STN CSCHEM, the usual search in CAS Registry, followed by crossover of the CAS RNs, retrieved only one labelled δ-ALA. A name search directly in STN CSCHEM yielded additional four labelled compounds; these were missed in the first search because no CAS RNs were assigned.

Pig liver esterase is tricky in CHEMCATS. The usual approach via CAS Registry retrieves one compound with 38 product records and 8 suppliers (for both STN CHEMCATS and SciFinder Scholar). A name search directly in STN CHEMCATS (not feasible in SciFinder Scholar) for those enzymes with no CAS Registry Number assigned is complicated by the different synonyms used (pig, porcine, hog), but gave one supplier record missed with the previous search. The same problems with name searching were encountered in ChemExper. Here, the desired compound could be found only with the search term „esterase“ among 3 compound records, but not with „pig liver esterase“. A CAS RN search directly lead to the desired compound. Three suppliers and 16 product records were listed.

Searches for chloromethyl polystyrene turned out to be a hard nut to crack. A molecular formula search in CAS Registry with (C9 H9 Cl)x was a good starting point for CHEMCATS in both SciFinder Scholar and STN. Three compounds were retrieved. This approach missed at least one such polymer that had no CAS RN assigned in CHEMCATS, but was retrievable by a search with name fragments. A structure search with variable chloromethyl positions at the benzene ring retrieved 2 commercially available compounds in SciFinder Scholar. ACD (DiscoveryGate) did well with name searching (8 compounds retrieved) and also provided the most information on products and suppliers (35). With ChemExper, different name variants had to be tried in sequence: 4 compounds were found with a total of 7 different suppliers. Structure searches for polymers did not work.

The not too uncommon reagent zinc borohydride – used in more than 10,000 reactions in the reaction database CASREACT – was neither found in CHEMCATS, CSCHEM, nor ChemExper.[35] However, a polymer-bound zinc borohydride available from Aldrich was retrieved in the ACD. A name search for commercially available “borohydride” retrieved 295 supplier records in STN CHEMCATS; the 265 of those which had CAS Registry Numbers assigned to them represent 18 different compounds. Among the unassigned borohydrides, we found five again on polymer support.

High-Tc superconducting lanthanum barium cuprates were not found in any of the catalogs. A search for high-Tc superconductors in a web directory, e.g. in Google directory [36], had more success and yielded several suppliers.

Commercial available zeolites are difficult to find because these inorganic solids comprise more than 160 different framework types, are not searchable by structure and have an uncounted number of different element compositions. Hence, only name search is possible, and name variants such as zeolite, molecular sieves, aluminosilicate and silicate must be considered. In STN CSCHEM, six records were found with a CAS RN search, based on a preceeding name search in CAS Registry with “zeolite” and “molecular sieve”, respectively. With “zeolite” searched directly in STN CSCHEM, four catalog records were retrieved, three of them missing from the first search because of lack of CAS RNs. A similar name search for “molecular sieve” in STN CSCHEM gave nine records, including seven not retrieved before. In addition to the 191 supplier records in STN CHEMCATS found with eleven different CAS RNs out of a preceeding name search in STN CAS Registry, we got nine zeolites (without assigned CAS RN) directly by name in STN CHEMCATS. An alternative search STN CA Registry with “molecular sieve” gave six CAS RNs, and corresponding 151 supplier records in STN CHEMCATS. Searching with “molecular sieve” directly in STN CHEMCATS retrieved a total of 297 supplier records; 141 of those had no CAS RN, and 47 of those had five different CAS RNs assigned. These were missed in the previous STN Registry name search because these compounds were named by CAS as silicic acid salts, not as molecular sieves.

In STN CSCHEM/CSCORP, the substructure search for taddoles was executed as usual in the CAS Registry file; of the total of 440 taddoles in CAS Registry, only five had records in CSCHEM (according to the CAS Registry Locator field). These taddoles were also covered by CHEMCATS (plus four more missing in STN CSCHEM). Two taddoles were only retrieved by a supplementary name search in CSCHEM itself, as these two “parent” taddole enantiomers from Lancaster Synthesis had no CAS RN assigned to them in CSCHEM.

The chiral phosphines compound class was used to test how chemical catalog databases can cope with stereo searches, since chirality may be brought in by stereogenic carbon as well phosphorus centers. We considered only phosphines in which the phosphorus atoms binds to three organic substituents, although one substituent may also be a hydrogen atom. In SciFinder Scholar, “chiral” was translated into a carbon atom with a stereobond to “any atom” to refine the many phosphines retrieved by a first substructure search in CAS Registry. In STN Registry, the files segment “stereosearch” was used instead. Both interfaces yielded the same number of compounds (83) for the fully substituted phosphines. For the single-hydrogen substituted phosphines, SciFinder Scholar found 40 commercially available compounds without stereo search, one of which was actually a stereo compound, but was missed after the stereochemistry refinement. In ACD (DiscoveryGate) and ChemExper, it was not possible to define the phosphorus atom as a stereogenic center in the MDL Draw and JME editor, respectively. ACD found only 4 fully substituted chiral phosphines, where chirality however was defined through stereogenic carbon atoms. Several compounds were missed in that search, as test searches in ACD with several CAS RNs of the chiral phosphines found in SciFinder Scholar showed.

The substructure search of tertiary alkyliodides requires to define generic groups („alkyl“). In STN CHEMCATS and CSCHEM this was achieved by a search in CAS Registry and transfer of the CAS RNs. The structure editors of the ACD (DiscoveryGate) and ChemACX did not provide generic groups as in CrossFire or SciFinder Scholar, e.g. an alkyl group. Therefore, iodides were retrieved which contained also other elements besides carbon and hydrogen.

Table 5 summarizes the results for the compounds discussed before and other searches not discussed because they were straightforward.

**Table 5.** Results of searches in chemicals catalog databases. Empty results: Search not done

| Compound | Search Type | Number of compounds / product records / suppliers | | | | | |
|---|---|---|---|---|---|---|---|
| | | CHEMCATS | | ChemACX | ACD | CSCHEM | ChemExper |
| | | SFS | STN | | | | |
| L(-)-glucose | Structure | 2 / 27 / 25 | | | 1 / 26 / 24 | | 0 / 0 / 0 |
| | CAS RN | | | | 1 / 26 / 24 | 1 / 4 / 19 | 0 / 0 / 0 |
| | Name | 1 / 26 / 25 | | | 1 / 26 / 24 | | 0 / 0 / 0 |
| Shikimic acid | Structure | 1 / 46 / 39 | | 1 / 1 / 1 | 1 / 39 / 36 | | 0[1] / 0 / 0 |
| | CAS RN | 1 / 46 / 39 | | | 1 / 39 / 36 | 1 / 3 / 27 | 1 / 41 / 37 |
| | Name | | | 2 / 20 / 17 | 1 / 39 / 36 | 1 / 3 / 27 | 1 / 41 / 37 |
| Estrone | Structure | 1 / 50 / 41 | | | 1 / 36 / 33 | 0 | 0[1] / 0 / 0 |
| | Incl. Isomers | 2 / 53 / 43[2] | | | 1 / 36 / 33 | | |
| | CAS RN | 1 / 50 / 41 | | | 1 / 36 / 33 | 1 / 4 / 28 | 1 / 33 / 28 |
| | Name | 1 / 50 / 41 | | | 1 / 36 / 33 | | 1 / 33 / 28 |
| Chloramphenicol | Structure | 1 / 87 / 47 | | | 1 / 76 / 52 | | 0[1] / 0 / 0 |
| | CAS RN | | | | 1 / 76 / 52 | 1 / 4 / 73 | |
| | Name | 1 / 87 / 47 | | | 1 / 76 / 52 | | 1 / 38 / 33 |
| Labelled δ-ALA | Structure (Exact) | | | | 0 / 0 / 0 | | 0 / 0 / 0 |
| | Incl.tautomers | | | | 8[3] | | |
| | Include salts | 2 / 2 / 2 | | | 0 / 0 / 0 | | |
| | CAS RN | | | | | 1 / 1 / 1 | 0 / 0 / 0 |
| | Name | | | | 7[3] | 7 / 7 / 5 | 0 / 0 / 0 |
| Pig liver esterase | Name | 1 / 38 / 8 | 1 / 38 / 8 | | 1 / 33 / 8 | 0 / 0 / 0 | 0 / 0 / 0 |
| | CAS RN | | 1 / 2 / 2 | | | 1 / 16 / 3 | 1 / 16 / 3 |
| Chloromethyl polystyrene | Structure | 2 / 20 / 4 | | | | | |
| | Name | 0 / 0 / 0 | | | 8 / 8 / 35 | 2 / 3 / 4 | 2 / 54 / 26 |
| | Molecular formula | 3 / 10 / 9[4] | 3 / 10 / 9[4] | | | 2 / 2 / 5 | |
| Polyethylenglycol 8000 | Name | 1 / 56 / 14 | | | 2 / - / 58 | 0 / 0 / 0 | |
| | Name (PEG) | 1 / 645 / 52 | 1 / 646 / - | | 2 / - / 58 | 1 / 26 / 98 | 1 / 234 / - |
| CAPSO | Structure | | | | 1 / 19 / 17 | | 1 / 25 / 22 |
| | Incl.tautomers | | | | 1 / 19 / 17 | | |
| | Include salts | 2 / 36 / 20 | | | 2 / 33 / 21 | | |
| | CAS RN | 1 / 24 / 17 | 2 / 33 /26 | | 2 / 33 / 21 | 2 / 4 / 13 | 1 / 25 / 22 |
| | Name | 1 / 24 / 17 | | | 1 / 19 / 17 | 2 / 2 / 13 | 1 / 25 / 22 |
| Zinc borohydride | Name | 0 / 0 / 0 | 0 / 0 / 0 | 0 / 0 / 0 | 1 / 1 / 1 | 0 / 0 / 0 | 0 / 0 /0 |
| Selectfluor | Name | 1 / 19 / 18 | 1 / 5 / 5 | | 1[5]/ 21 / 18 | 0 / 0 / 0 | 1 / 17 / 16 |
| | CAS RN | 1 / 19 / 18 | | | 1 / 21 / 18 | 1 / 2 / 12 | 1 / 17 / 16 |
| $Yb(OTf)_3$ | Structure | | | | 2 / 25 / 19 | | 0 / 0 / 0 |
| | CAS RN | 1 / 25 / 21 | | | 2 / 25 / 19 | 1 / 3 / 16 | 1 / 18 / 15 |
| | Name | 1 / 25 / 21 | | | | | 0 / 0 / 0 |
| TiN | Structure | 1 / 27 / 14 | | | | | - |
| | CAS RN | | 1 / 27 / 14 | | 1 / 32 / 17 | 1 / 1 / 18 | |
| | Name | | 1 / 30 / 16 | | 1 / 32 / 17 | 1 / 1 / 18 | 1 / 16 / 11 |
| | Molecular formula | 1 / 27 / 14 | | | | | 1 / 20 / 11 |
| Zeolites | Name | -[6] | -[6] | -[6] | -[6] | -[6] | -[6] |
| Taddoles[3] | Substructure | 11 | 9 | 7 | 15 | | 2 |
| Chiral phosphines[3] | Substructure | 83 | 83 | | | 55 | -[7] |
| Tert. alkyliodides[3] | Substructure | 3 | 3 | | 6 | | 2 |
| La-Ba cuprates[3] | Composition | 0 | 0 | 0 | 0 | 0 | 0 |
| Disubstituted ferrocenes[3] | Substructure | 6 | | | 3 | | -[8] |

[1] compound only found with substructure search

[2] two records of estrone with absolute (CAS RN 57-16-7) and relative (CAS RN 19973-76-3) configuration

[3] only number of found compounds given

[4] overlap one only!

[5] name fragment only

[6] see text

[7] not searchable, gave error in stereoprocessing (1640 phosphine and phosphonium compounds, mostly not chiral)

[8] not searchable

# 4 Discussion

## 4.1 Search Facilities

All catalog databases must be searchable by full or partial structure, as this is normally the easiest and most precise way of searching for commercial available compounds. This condition is fulfilled for all major sources, albeit for CHEMCATS and STN CSCHEM not directly, but only via (sub)structure searches in CAS Registry. Substructure searches are straightforward in all commercial systems examined. Searches for exact structure, however, demand some knowledge about the features of the system. With respect to this, STN Messenger is most demanding, but also offers the most precise and powerful way of searching. SciFinder (Scholar) always does broad searches, e.g., the pyranose structure of glucose retrieves both this and the open aldose form as well, with "Analyze by Precision" to refine if desired. MDL DiscoveryGate and ChemACX need a more closer look, as exact searches retrieve just the entered structure, missing out on stereoisomers (or a record with stereochemistry missing), tautomers, salts, and labelled compounds. We strongly recommend to use "Full Structure" in ChemACX, and "Include Tautomers" plus a second search in DiscoveryGate with the same structure and "Include Salts" when appropriate. We found that the option "Include Tautomers" includes also the labelled compounds; this should be documented in the help messages.

Although chemical name searching is problematic because of lack of standardization and lack of adherence to strict nomenclatures of supplier-provided names, it is important for compounds with no well-defined structures, such as many inorganics, materials, and polymers, or with no structure present in the database (enzymes and other biopolymers). SciFinder (Scholar) had some problems with polymers because CAS Registry entries were not consistently named (poly-something vs. something polymer). In addition, searching of name fragments is not possible within SciFinder (Scholar). In ACD under DiscoveryGate, name searches for name fragments ("contains") were very slow. ChemExper offers truncation in Advanced Search mode („IUPAC value" and „contains"), but is slow as well.

Given the problems with chemical nomenclature in general, and supplier names in catalogs in particular, printed catalogs wer preferentially searched with molecular formulas. With this in mind, it comes as a surprise that of the major catalogs examined, only ACD, ChemACX, and ChemExper, but not CHEMCATS and STN CSCHEM, offer this search facility. In the latter two, searches for exact molecular formulas, or more generally, for composition of compounds - e.g., all compounds that contain Ti and N in any relation, but no other element - are only possible via a corresponding search in the CAS Registry compound database (SciFinder (Scholar) for CHEMCATS, STN Messenger for CHEMCATS and CSCHEM), followed by crossover of the CAS RNs thus retrieved into the catalog database. SciFinder (Scholar) permits only searches for exact molecular formulas, not for partial ones or elemental composition. In ChemExper, molecular formulas are range-searchable, and can be entered without a prescribed order of the elements. It even recognizes common abbreviations for substituents (such as Me, Et, Tf, etc.) and translates them.

As unique and common identifiers for chemical substances, CAS RNs had already played a substantial part in printed chemicals catalogs. Their significance has increased with the availability of corresponding catalog databases. CAS RNs are not only precise access points within the catalog databases, they are also a link to the largest compound and structure database, the CAS Registry system. This gives users the full searching power in this database, under SciFinder (Scholar) and even more so under STN Messenger. This is particularly important with respect to the somewhat limited search facilities in the catalogs themselves: supplier-provided name, structure, and molecular formula search are available in ACD or ChemExper, but not in STN CSCHEM and CHEMCATS, with the exception of names. There is a caveat, however: Searching by CAS RN is critically dependent on complete and correct assignment of these numbers in catalog databases. Both according to our general experience and with respect to the examples studied, this cannot be taken for granted. Simple organic compounds are less of a problem in this respect, but enzymes and other biopolymers are, as well as materials. For such queries, as in our searches for estrone, chloromethyl polystyrene, taddols and δ-aminolevulinic acid, a RN search must be supplemented by name and molecular formula search.

Bulk suppliers are flagged out in DiscoveryGate ACD with the flag B in the data field “bulk_fine”, and with BLK in STN CSCHEM. Both flags can be displayed (in the case of ACD, they can also be exported into a spreadsheet), but to our knowledge, they are directly searchable to limit to such suppliers only in STN CSCHEM.

### 4.2 User Interfaces

Databases of chemicals directories are accessed either via WWW browsers, with dedicated client software, or with retrieval languages at hosts. For ease of use, WWW-based solutions seem to be the best option. On one hand, they offer easy navigation, search history tracking (at least by way of the back and forward browser buttons), easy print-out of pages, independence of hardware and operating system used, links to additional documentation (e.g. MSDS or primary literature), physical property data and supplier web sites, and interoperability such as the possibility to send requests for offers or orders to the suppliers.

On the other hand, browser compatibility problems and browser plugins or Java applets which are required for structure input and display do cause problems or narrow the availability for different operating systems. Unless they are nearly completely Java-based such as DiscoveryGate or Web 2.0 enabled such as eMolecules, WWW-based chemicals directories do most often not have the functionality and responsiveness compared to desktop applications. They work in the scheme create and submit query – wait for results (and do nothing in the wait time) – display first set of results – request another set of results if needed – wait – display and so on. Within desktop applications such as SciFinder (Scholar), the user may interrupt these processes, or copy a CAS RN to the clipboard and paste the corresponding structure in the structure editor for further editing. Retrieval-language based systems allow to execute automated batch scripts.

Most of the GUIs of the examined WWW and client-software based chemical catalogs are well designed and help the users to achieve their jobs. In our opininon, MDL DiscoveryGate offers the most feature-rich and flexible GUI with respect to both input of queries (both structure and properties) and display of results. However, sorting of compounds on given criteria did not work in the Mac OS X version. ChemACX uses the ChemDraw plugin for structure input. This editor is very powerful, but many of the functions such as Undo are not available in the tool palette, but by clicking on the right mouse button only. Whereas the SciFinder (Scholar) structure editor is very user-friendly, the display of CHEMCATS results is a dead-end. For price and supplier address information, every product record in the list of products must be clicked seperately. Comparison of product information is hampered because the window content is always replaced with new information. Supplier URLs are given, but do not function as links. eMolecules offers a very flexible search and display mode in its expert search and comes closest to Web 2.0 applications. ChemExper provides a wealth of additional information such as physical data, links to MSDS, IR and NMR spectra, and 3D models. However, record navigation both back to the list of compound hits and to the previous or next record is missing.

### 4.3 Data structure

We observed for all chemicals directories one major drawback: Product, supplier, package, price and other data are not normalized in the sense of relational database theory. This is mainly due to the fact that the chemical information producers (or aggregators) do not obtain data from the suppliers in a standardized way, but also because there is not much effort to clean up the obtained data. This first has consequences for sorting. Most of the databases, if ever, allow sorting for a few fields only, such as supplier name or order number, but not for prices, package size, or purity of the chemical. LabVelocity was the only notable exception where products can be sorted by price. Second, it hampers post-processing of exported records, e.g., in spreadsheets, lab inventory systems and databases. Third, it prevents functions that provide added value, such as calculation of prices per 100 g or currency conversion, or interoperability with the (web-)shops of the suppliers. Last, it hampered our comparison of databases with respect to assignment of product records.

### 4.4 Product prices

Product prices are often not provided. If they are available, they must be taken with caution, because shipping and packing charges and taxes may be included or not, dependent on the supplier. Comparison of prices is further hampered by the fact that the same product may have different prices in different countries. At least, in online databases on a pay-per-use basis (STN CHEMCATS, STN CSCHEM), product records without price information should not be charged for!

### 4.5 Database integration

Catalog databases are, even more than other chemical databases, part of systems and processes for chemical information retrieval. They are particularly related, and thus need to be closely integrated with, reaction databases and general compound databases. SciFinder (Scholar), STN Messenger, and MDL DiscoveryGate as well as the older MDL ISIS/Base provide this kind of integration. Second, chemical catalogs can also be part of systems for local management of chemical information, e.g., ChemACX and CambridgeSoft’s ChemOffice Webserver to provide integration with enterprise inventory systems. Similarly, the MDL Isentris suite offers integration and management of enterprise data with MDL databases such as ISIS/ACD, as well as external databases. Expereact [13], a chemical inventory system used at ETH Zürich and provided by ChemExper, is augmented with supplier-provided data. Web-based catalogs are lacking both types of integration to some respect, but probably will catch up by providing web services which enable integration.

### 4.6 Import facilities

Sometimes, one needs not only individual searches for compounds in catalogs, but import of many query terms (structures, CAS RNs, chemical names) into catalog databases to execute automated searches for a large number of substances. In our opinion, none of the major commercial catalog databases provides at present such facilities in a sufficient way. Only STN Messenger via the front-end software STN permits import of a large number of textual queries via command files. In SciFinder (Scholar), import is limited to 25 terms, which is clearly insufficient. MDL DiscoveryGate allows to import large lists of their internal MDL catalog numbers, but these in return must first be generated by a search in DiscoveryGate itself! Import of single structures is possible via the structure editor in most of the structure-searchable chemicals catalogs, except where the JME editor is used. According to our experience, import of a batch of structures is only possible with the free database ZINC, into which lists of SMILES strings can be uploaded.

## 4.7 Export facilities

Printing to Adobe PDF (all systems) as well as exporting to HTML (MDL DiscoveryGate) or RTF (SciFinder) is useful for reading and documentation, but practically useless for any further post-processing. From SciFinder (Scholar), data can post-processed best if saved in quoted (i.e. comma-separated) or tagged format, the latter needing a parser to be written. Catalog databases under the command-driven system STN Messenger (CHEMCATS, STN CSCHEM) offer printing of text and structures via conversion of capture files into RTF in a very flexible way, particularly when using the ANALYZE command for selective extraction of fields (compound names, CAS RNs, supplier names, etc.). The resulting ASCII files, however, are not amenable to post-processing without extensive manual or scripted editing. The best options in that respect are exporting to MS Excel spreadsheets as in MDL ISIS/Base, DiscoveryGate or eMolecules. Other options, like export to SDF files (ChemACX, MDL Discovery Gate) are considered too complex and unwieldy for end-users. Overall, MDL ACD and eMolecules in its Expert mode offer the most flexible export facilities of the systems and databases examined. As discussed before, post-processing of the exported data is hampered by the fact that data is not normalized.

# 5 Conclusions

For the average end-user, web-based chemicals catalogs are in our opinion the most prefered solution. Although in many cases the free catalogs presented here do not provide the same coverage of products than their commercial equivalents, they will catch up quickly. For example, since its release one year ago, eMolecules has grown to one of the largest aggregators of commercially available chemicals, although it is confined to compounds with structural data. The investigated web-based systems showed great potential of what functions can be possible by integrating and linking to external information, such as MSDS, safety data, spectra, 3D model data, literature, or deep links to supplier databases. We believe that this trend will go further when catalog systems on the WWW will provide in an integrated fashion sorting, price comparison of products, news and alerts on new and updated products, Amazon-like one-stop web-shopping, and start to use web services and mashups, e.g. for on-time querying of currencies from the next bank to calculate current prices, tracking of shippings, or using localization services.

In industry, chemicals catalog databases that can be integrated into chemical information management system will have a clear advantage over closed systems. At present, this is only available with dedicated client software, e.g., where ChemACX fits in CambridgeSoft's ChemOffice, ACD in MDL's ISIS and ISENTRIS, and CAP in Accelrys Accord, respectively. Web-based systems may be enabled to this type of integration if they will provide web services like SOAP [37] for open exchange of information.

Structure searching is still most powerful in the commercial systems, the best and easiest system in our opinion being SciFinder (Scholar), followed by ACD (DiscoveryGate). DiscoveryGate however is better with respect to display and exporting of results, whereas SciFinder (Scholar) needs improvement. As best practice for searching we recommend first structure searches (which in STN must be carried out in CAS Registry involving subsequent transfer of CAS RNs to CHEMCATS or CSCHEM), second searching by name, although several databases show problems with name fragments.

## 6 Acknowledgements

We thank CambridgeSoft Corp. and Elsevier MDL for providing us test access and support for their ChemACX Enterprise and MDL DiscoveryGate products, respectively. We thank also Dr. Luc Patiny from EPFL for his useful comments.

## 7 References

[1] Amazon.com: Online Shopping for Electronics, Apparel, Computers, Books, DVDs & more, http://www.amazon.com/ (last access 16.10.2006)

[2] CHEMCATS database, http://www.cas.org/CASFILES/chemcats.html (last access 14.10.2006)

[3] CHEMCATS Database Summary Sheet, http://www.cas.org/ONLINE/DBSS/chemcatsss.html (last access 14.10.2006)

[4] Murray D. Rosenberg, Marian Z. DeBardeleben, John F. DeBardeleben, J. Chem. Inf. Comput. Sci. 1982, 22(2), 93.

[5] Elsevier MDL :: Products :: MDL® Discovery Knowledge :: DiscoveryGate®, http://www.mdl.com/products/knowledge/discoverygate/ (last access 14.10.2006)

[6] Databases in Science and Technology - STN International, http://www.stn-international.com/ (last access 14.10.2006)

[7] CSCHEM, http://www.stn-international.com/stndatabases/databases/cschem.html (last access 14.10.2006)

[8] CSCHEM Database Summary Sheet, http://www.cas.org/ONLINE/DBSS/cschemss.html (last access 14.10.2006)

[9] CSCORP, http://www.stn-international.com/stndatabases/databases/cscorp.html (last access 14.10.2006)

[10] CSCORP Database Summary Sheet, http://www.cas.org/ONLINE/DBSS/cscorpss.html (last access 14.10.2006)

[11] Chem Sources - Chemical Purchasing Directory Buyers Guide, http://www.chemsources.com/ (last access 14.10.2006)

[12] Accelrys: Chemical Databases: Chemicals Available for Purchase, http://www.accelrys.com/products/chem_databases/databases/CAP.html (last access 16.10.2006)

[13] ChemExper - catalog of chemical suppliers, physical characteristics and search engine, http://www.chemexper.com/ (last access 14.10.2006)

[14] Jmol, open source molecule viewer written in Java, http://jmol.sourceforge.net/ (last access 14.10.2006)

[15] Google Advertising, http://www.google.com/ads/ (last access 14.10.2006)

[16] JME Molecular Editor, by Peter Ertl, Novartis. http://www.molinspiration.com/jme/ (last access 14.10.2006)

[17] ChemNet, http://www.chemnet.com/ (last access 14.10.2006)

[18] eMolecules Chemical Search, http://www.emolecules.com/ (last access 14.10.2006)

[19] CambridgeSoft Corp., ChemDraw, http://www.cambridgesoft.com/software/ChemDraw/ (last access 14.10.2006)

[20] Elsevier MDL :: Products :: MDL® Discovery Framework :: MDL® ISIS/Draw, http://www.mdl.com/products/framework/isis_draw/ (last access 14.10.2006)

[21] ICISsearch, http://www.icissearch.com/ (last access 14.10.2006)

[22] ResearchLink by LabVelocity, http://researchlink.labvelocity.com/ (last access 14.10.2006)

[23] Super Natural Database, http://bioinformatics.charite.de/supernatural/ (last access 14.10.2006)

[24] Accelrys Catalyst, http://www.accelrys.com/products/catalyst/ (last access 14.10.2006)

[25] MarvinSketch & MarvinView, http://www.chemaxon.com/marvin/ (last access 14.10.2006)

[26] Elsevier MDL :: Products :: MDL® Discovery Framework :: MDL® Chime, http://www.mdl.com/chime/ (last access 14.10.2006)

[27] ZINC – a free database for virtual screening, http://blaster.docking.org/zinc/ (last access 14.10.2006)

[28] John J. Irwin, Brian K. Shoichet, J. Chem. Inf. Model. 2005, 45(1), 177. DOI: 10.1021/ci049714+ (http://dx.doi.org/10.1021/ci049714+)

[29] Daylight Chemical Information Systems Inc., SMILES$^{TM}$, Simplified Molecular Input Line Entry System, http://www.daylight.com/smiles/ (last access 14.10.2006)

[30] The PubChem Project, http://pubchem.ncbi.nlm.nih.gov/ (last access 14.10.2006)

[31] Google Directory: Business > Chemicals, http://www.google.com/Top/Business/Chemicals/ (last access 14.10.2006)

[32] DMOZ Open Directory: Business > Chemicals, http://www.dmoz.org/Business/Chemicals/ (last access 14.10.2006)

[33] Chemie-Marktplätze im Internet, http://www.nrwchemie.de/wir/links_marktplatz.htm (last access 14.10.2006)

[34] K. Weissermel, H.-J. Arpe, Industrial Organic Chemistry, 4th ed., Wiley-VCH, Weinheim, p. 277 ff. (2003).

[35] This is consistent with observations reported to us. A preparation procedure is described on the ORGLIST archive, http://www.orglist.net/archive/2000/0252.html (last access 14.10.2006)

[36] Google Directory, Science > Technology > Materials > Superconductors, http://www.google.com/Top/Science/Technology/Materials/Superconductors/ (last access 14.10.2006)

[37] World Wide Web Consortium, Simple Object Access Protocol, http://www.w3.org/TR/soap/ (last access 14.10.2006)